Genomic organization underlying deletional robustness in bacterial metabolic systems

Hosseini, Sayed-Rzgar ; Wagner, Andreas

Abstract: Large-scale DNA deletions and gene loss are pervasive in bacterial genomes. This observation raises the possibility that evolutionary adaptation has altered bacterial genome organization to increase its robustness to large-scale tandem gene deletions. To find out, we systematically analyzed 55 bacterial genome-scale metabolisms and showed that metabolic gene ordering renders an organism's viability in multiple nutrient environments significantly more robust against tandem multigene deletions than expected by chance. This excess robustness is caused by multiple factors, which include the clustering of essential metabolic genes, a greater-than-expected distance of synthetically lethal metabolic gene pairs, and the clustering of nonessential metabolic genes. By computationally creating minimal genomes, we show that a nonadaptive origin of such clustering could in principle arise as a passive byproduct of bacterial genome growth. However, because genome randomization forces such as translocation and inversion would eventually erode such clustering, adaptive processes are necessary to sustain it. We provide evidence suggesting that this organization might result from adaptation to ongoing gene deletions, and from selective advantages associated with coregulating functionally related genes. Horizontal gene transfer in the presence of gene deletions contributes to sustaining the clustering of essential genes. In sum, our observations suggest that the genome organization of bacteria is driven by adaptive processes that provide phenotypic robustness in response to large-scale gene deletions. This robustness may be especially important for bacterial populations that take advantage of gene loss to adapt to new environments.

DOI: https://doi.org/10.1073/pnas.1717243115

PNAS
www.pnas.org

Supplementary Information for

Genomic organization underlying deletional robustness in bacterial metabolic systems

Sayed-Rzgar Hosseini, and Andreas Wagner

Corresponding authors:

Sayed-Rzgar Hosseini (Email: razgar@gmail.com)

Andreas Wagner (Email: andreas.wagner@ieu.uzh.ch)

This PDF file includes:

Supplementary texts S1 to S13
Figs. S1 to S48
Tables S1 to S27
References for SI reference citations

Supplementary texts:

Text S1: The DNA deletion rate exerts sufficient selection pressure on bacterial genomes to be visible to natural selection.

Previous work has shown that natural selection can increase robustness to any one kind of genetic mutation when the product of effective population size N_e and mutation rate u exceeds one ($N_e u >> 1$), essentially because under these conditions a population will be polymorphic for the variants created by the mutation, and selection requires polymorphism (1).

Thanks to mutation-accumulation experiments and whole-genome sequencing, estimated rates of insertion and deletion (indel) mutations (u_{id}) are available for multiple species (2). Prokaryotic genomes show a deletion bias (3), i.e., the rate of DNA deletion is higher than that of insertion, which implies that half the rate of indels ($u_{id}/2$) can serve as a lower bound for the deletion rate per site, from which one can calculate the deletion rate per genome (U_{del}) as the product of $u_{id}/2$ with the effective genome size. Table S1 below is adapted from Table 1 in Sung et al (2), and summarizes the available information on the deletion rate for seven bacterial species.

According to this data, $N_e U_{del} >> 1$ for all seven species, implying that gene deletion can exert strong enough selection pressure on the bacterial genomes in question. Additional data comes from experimental evolution of *Salmonella enterica*, where a combination of sequential bottlenecking of colonies and pulsed-field gel electrophoresis was used to measure the rate of large-scale deletions (4). This study revealed gene deletions ranging from 1kb to 202 kb in size, which occurred at a high rate of approximately 0.02 per genome and generation which also implies that $N_e U_{del} >> 1$ given typical bacterial population sizes. In sum, empirical data support the notion that the frequency of gene deletions is sufficiently high that selection for deletional robustness can be effective.

Species	Reference	U_{del}($\times 10^{-3}$ per generation)	N_e ($\times 10^{6}$)	N_e U_{del}
Agrobacterium tumefaciens	Sung et al. (2)	0.075	342.47	25,685.25
Bacillus subtilis	Sung et al. (5)	0.216	61.19	13,217.04
Escherichia coli	Lee et al. (6)	0.07215	179.60	12,958.14
Mesoplasma florum	Sung et al. (7)	0.8085	1.07	865.09
Pseudomonas aeruginosa	Sung et al. (8)	0.0413	210.70	8701.91
Staphylococcus epidermidis	Sung et al. (2)	0.1186	35.14	4,169.36
Vibrio cholerae	Sung et al. (2)	0.0306	478.26	14,634.75

Table S1. Lower bounds for population deletion rates in bacterial species. Each row contains data from the indicated bacterial species. Columns, from left to right, show species or strain names, a literature reference, deletion rate per generation (U_{del} ($\times 10^{-3}$)), effective population size (N_e ($\times 10^{6}$)), and N_e U_{del}.

Text S2: Quantifying robustness based on kilo base pairs (kbp) of deleted DNA instead of the number of deleted metabolic genes.

In this analysis, we used an alternative approach to quantify the robustness to tandem and random deletions, which is not based on the number of deleted metabolic genes, but on the amount of contiguous DNA deleted in kbp. This approach takes into account that metabolic genes do not have equal length and need not be equally spaced and uniformly distributed in a genome. To make the results of this approach comparable across genomes, we defined a unit of DNA deletion for any given genome as the genome length in which one metabolic gene is expected to be located, on average. We determined this unit length by dividing the total length of a genome by the sum of the length of metabolic genes in the genome. It computed as 3.290, 5.240 and 3.640 kbp, respectively, for *Escherichia coli K-12 MG1655 (iJO1366), Bacillus subtillis* and *Salmonella enterica*. Then, we divided each genome into non-overlapping segments of the unit length. To analyze the effects of tandemly deleting *k* length units, we deleted all possible consecutive DNA segments comprising *k* units. For random deletion, we generated the same number of deletional variants, but deleted *k* randomly selected segments of unit length. As in our main previous approach, we varied *k* from 1 to 50. Figures S4 and S5 show that the main observations remain similar to those of our main approach.

Text S3: Essential metabolic genes are clustered in bacterial genomes

We hypothesized that the clustering of essential genes can increase the robustness of the genome to simultaneous deletion of multiple successive (*tandem*) genes. In contrast, it should not impact robustness to simultaneous deletions of multiple genes, chosen at random, regardless of their genomic location. If so, the clustering of essential genes might help explain the excess robustness to tandem deletions. To find out whether this is the case, we first identified all essential metabolic genes in each of our 55 study genomes and in each of the 102 minimal environments (Table S3). We then used the test statistic of Kuiper's test (Methods) to calculate a measure of clustering, i.e., the extent to which the distribution of essential genes differs from a uniform distribution. In this analysis we distinguished two types of essential genes: i) strictly essential metabolic genes, which are essential on all carbon sources, and ii) conditionally essential metabolic genes, which are essential on at least one carbon source (see methods). In the vast majority of the species, both types of essential genes are significantly clustered in the genome (Tables S4 and S5). What is more, we observed a positive correlation between the degree of essential gene clustering and robustness to tandem gene deletions (deletion length 5: Pearson's *r=0.26*, and *P=0.05;* length 10; *r=0.35*, and *P =0.009*)), but no significant correlation with robustness to random deletions (of length 5

(Pearson's *r=-0.04*, and *P-value=0.77*) and of length 10 (Pearson's *r=-0.01*, and *P-value=0.95*)). This confirms the importance of essential gene clustering for robustness to tandem gene deletions.

Text S4: Long non-essential clusters of non-essential genes

We observed that bacterial genomes harbor long clusters of non-essential metabolic genes that are not interrupted by any essential genes. If simultaneous deletion of all these non-essential genes does not abolish viability, then this arrangement in itself increases the robustness of bacterial genomes to tandem deletions. We distinguish between two different types of non-essential metabolic genes: *i)* strictly non-essential genes that are not essential on any carbon sources, and *ii)* conditionally non-essential genes that are not essential on at least one carbon source.

Any two successive strictly essential metabolic genes are either adjacent in the metabolic genome or non-adjacent, i.e., separated by one, two, or a larger cluster of non-essential metabolic genes. The non-essential genes belonging to such a cluster are at least conditionally non-essential, meaning that they are non-essential on at least one carbon source, but they are not necessarily strictly non-essential. Thus, we call a cluster of non-essential metabolic genes intervening between two successive strictly essential metabolic genes a "*cluster of conditionally non-essential metabolic genes".* After identifying all clusters of conditionally non-essential metabolic genes, we aimed to determine whether each such cluster is essential for viability on the carbon sources we consider. To do so, we deleted all metabolic genes in a given cluster, translated the gene deletions into reaction deletions, and used FBA to determine the resulting metabolism's viability on the set of carbon sources on which the wild-type metabolism (before deletion) was viable. If the deletion did not abolish viability on any of these carbon sources, we called such a cluster a "*strictly non-essential cluster of conditionally non-essential metabolic genes*". Moreover, if the deletion did not abolish viability on at least one carbon source on which the wild-type metabolism was viable, we called the cluster a "*conditionally non-essential cluster of conditionally non-essential metabolic genes*". Note that the set of strictly non-essential clusters (of conditionally non-essential metabolic genes) is a subset of the set of conditionally non-essential clusters (of conditionally non-essential metabolic genes) (See figure S14).

We can apply a similar procedure for any cluster of *strictly non-essential* metabolic genes intervening between two successive (but not adjacent) conditionally essential metabolic genes. In this way, we can identify *strictly non-essential clusters of strictly non-essential metabolic genes* and *conditionally non-essential clusters of strictly non-essential metabolic genes*. Note that again the set of strictly non-essential clusters (of strictly non-essential metabolic genes) are

a subset of the set of conditionally non-essential clusters (of strictly non-essential metabolic genes). However, the set of clusters of strictly non-essential metabolic genes are not a subset of the set of the clusters of conditionally non-essential metabolic genes (See figure S14).

We observed that both strictly and conditionally non-essential genes are organized into fewer but longer clusters in wild-type genomes than in randomized genomes (Figure S15). For example, in *Escherichia coli K-12 MG1655 (iJO1366)*, we observed a long cluster of 32 consecutive strictly non-essential metabolic genes and a cluster of 65 consecutive conditionally non-essential metabolic genes. We observed that simultaneous deletion of all the genes in a given cluster of strictly non-essential metabolic genes does not abolish viability on any carbon sources in more than 95% of the clusters, and it does not abolish viability on at least one carbon source in more than 99% of the clusters for the vast majority of the genomes (50 out of 55, that is 90.9%) (See table S10). Moreover, in around 90% of the clusters of conditionally non-essential metabolic genes, simultaneous deletion of all the genes belonging to the cluster does not abolish viability on at least one carbon source, and in around 50% of these clusters in all genomes, it does not abolish viability on any carbon source (See table S11). This carbon-source dependent viability loss caused by deletion of the clusters of non-essential genes can explain the variability among different carbon sources that we observed (Figures S8 and S9).

Text S5: Passive emergence of the clusters of essential genes

We hypothesized that an evolutionary genome expansion scenario could explain essential gene clustering in present-day bacterial genomes. To validate this hypothesis, we started from minimal genomes(9), that is, sets of metabolic genes from which not a single gene can be removed without abolishing viability on a specific carbon source (See methods in the main text). Specifically, we started with 100 distinct minimal metabolic genomes, which we had derived from the *E. coli K12* wild type genome through stepwise elimination of genes that are nonessential for viability on glucose as the sole carbon source. Starting from one of these minimal genomes, we then chose randomly (with a uniform distribution) between one and five not necessarily contiguous metabolic genes from the wild type *E. coli K12* genome that are not already included in the minimal genome. We then inserted the selected genes as a contiguous set of genes into a randomly chosen position in the genome. This procedure implies that during an insertion event on average three genes are added to the genome. This choice of including multiple genes in an insertion event is motivated by ample empirical evidence in favor of co-acquisition and co-insertion of multiple genes into bacterial genomes, for example during horizontal gene transfer events(10–14). We repeated this insertion process, taking care to only choose genes for insertion that were not already present in the

growing recipient genome, until our genome had reached the size of the wild type *E. coli K12* genome. We note that the gene content of the resulting genome is identical to that of the wild type genome, but its gene order is not. We repeated this procedure for all 100 starting minimal genomes. We observed that the robustness of the resulting genomes to tandem gene deletion (figure S16a, blue boxes) is considerably higher than that of a randomly reshuffled *E.coli K-12* genome (figure S16a, black boxes), and comparable to that of the *E.coli K-12* wild-type genome (figure S16a (blue horizontal line)). Because the final genomes produced by these simulations have the same set of genes as that of *E. coli K-12*, the fraction of essential genes is exactly the same as that of *E.coli K-1*2 (Figure S16b). A difference in this fraction to *E.coli K-1*2 can thus not possibly be responsible for the increase in robustness (figure S16b). In the majority of the resulting genomes, essential genes are also significantly clustered (figure S16c). Thus, essential genes can passively get clustered in the genome as a byproduct of genome growth processes.

We then examined additional evolutionary forces, such as gene deletion and duplication, which might further enhance robustness to tandem deletions. More specifically, we compared four different kinds of genome-altering changes to expand each of 100 minimal genomes that are viable on glucose and derived from the *E.coli K-12* genome to a size that is identical to that of the *E.coli* K-12 genome. That is, we expanded genomes through: i) insertion events alone (as just described), ii) insertion and deletion events, iii) insertion and duplication events, and iv) insertion, duplication, and deletion events, as described below:

Insertion + deletion: We started with 100 distinct minimal genomes derived from the *E.coli K-12* genome and viable on glucose. For each of these genomes, we performed the following iterative procedure. In each step of this procedure, we either i) (with 0.75 probability) randomly and with a uniform distribution chose between 1 to 5 (not necessarily contiguous) metabolic genes from the set of genes that were present in the full-size genome, but absent from the minimal genome, and inserted them as a contiguous gene cluster at a randomly chosen position in the recipient genome, or ii) (with 0.25 probability) deleted a randomly chosen non-essential cluster of genes in the growing genome. We iterated this procedure until the growing genome had reached a size equal to that of the focal species. Note that this procedure also avoids duplication of genes in the growing genome, and ensures that all genes in the full-size wild type genome will be included in the final genomes. The higher insertion than deletion probability ensures that genome size grows over time. Note that including deletions implies that more steps are needed to reach the final genome size.

Insertion + duplication: We started with 100 distinct minimal genomes derived from the *E.coli K-12* genome and viable on glucose. For each of these genomes, we performed the following iterative procedure. In each step of this procedure, we either i) (with 0.75

probability) randomly and with a uniform distribution chose between 1 to 5 (not necessarily contiguous) metabolic genes from the set of genes that were present in the full-size genome, but absent from the minimal genome, and inserted them as a contiguous gene cluster at a randomly chosen position in the recipient genome, or ii) (with 0.25 probability) we duplicated a given number of genes (between 1 to 5) chosen at random. We iterated this procedure until the growing genome had reached a size equal to that of the focal species. In these simulations, gene duplications contributed to 0.25 of the added genes, which implies that the final genome will not contain some of the genes in the wild-type genome of the focal species.

Insertion + duplication + deletion: We started with 100 distinct minimal genomes derived from the *E.coli K-12* genome and viable on glucose. For each of these genomes, we performed the following iterative procedure. In each step of this procedure, we either i) (with 50% probability) randomly and uniformly chose between 1 to 5 (not necessarily contiguous) metabolic genes from the set of genes that were present in the full-size genome, but absent from the minimal genome, and inserted them as a contiguous gene cluster at a randomly chosen position in the recipient genome, or ii) (with 0.25 probability) duplicated a given number of genes (between 1 to 5) chosen at random, or iii) (with 0.25 probability) deleted a randomly chosen non-essential cluster of genes in the growing genome. We iterated this procedure until the growing genome had reached a size equal to that of the focal species.

We note that in simulations including deletion, the insertion probability during the genome growth procedure must be larger than the deletion probability. Otherwise, a genome will not grow in size. Moreover, to make the results of different scenarios above comparable, we kept the deletion and duplication probabilities at the same value of 0.25. At the end of each simulation, and for each of the 100 final full-sized genomes, we quantified metabolic robustness to tandem deletions of five genes, identified essential genes in each genome, and studied whether these genes are clustered. In estimating robustness to multi-gene deletions, we focused on deletions of five genes. However, we repeated our analysis with deletions of n=10 genes, to find out whether the patterns we observed were insensitive to changes in deletion size. Computational cost prohibited a more exhaustive exploration of deletion sizes in these simulations.

We observed that most of the genomes subject to both insertion and deletion reached higher final robustness to tandem deletions than for the other three scenarios, and even higher robustness than wild-type *E.coli* (figure S16a (cyan boxes)). Thus, exposing our simulated genomes to tandem deletions leads to higher robustness to such deletions. Including duplications slightly lowers robustness compared to the insertion-only scenario. To understand why, consider the following. In the absence of gene duplication, the final genome

contains the same genes as that of *E.coli K12*, albeit with different order. In the presence of duplication, however, some *E.coli K12* genes may have multiple copies in the final genome, whereas others may be missing. The fraction of essential genes is the same as that of *E.coli K12* in the absence of duplication, but it is smaller in the presence of duplication, because duplication of an essential gene results in two individually non-essential genes (figure S16b). Furthermore, in the presence of gene duplication, essential genes do not become significantly clustered, whereas under insertion and deletion alone, essential genes in the second scenario are more strongly clustered than *E.coli K12* (figure S16c). Because gene duplication disrupts clusters of essential genes, it does not increase robustness to tandem gene deletion despite lowering the fraction of individually essential genes.

To check whether the ordering of the genes in the initial minimal genomes is important for the patterns we observed, we repeated the above procedure starting with minimal genomes in which the ordering of the metabolic genes is reshuffled. Using this approach we observed the same patterns (figure S17). Moreover, our observations remain the same when we use acetate instead of glucose as the sole carbon source in the minimal medium (figure S18). Although our qualitative observations are not strongly sensitive to parameters such as the insertion to deletion or duplication ratios (figure S19), increasing the number of inserted genes, increasing the deletion probability, and decreasing the duplication probability systematically increases the clustering of essential genes and the robustness to tandem deletion (figure S19). The observed patterns remain unchanged when we quantify (larger-scale) deletional robustness by deleting 10 instead of 5 metabolic genes (figure S20). Finally, we get similar results when we repeat this procedure using other genomes (figure S21).

In sum, we conclude that clusters of non-essential genes observed in wild-type genomes might originate non-adaptively. In other words, they can emerge passively through sequential insertion of new genes into a minimal genome. Moreover, exposure to gene deletion enhances clustering of essential genes, but duplication can disrupt clusters of essential genes.

Text S6: Genome rearrangement can disrupt essential gene clusters

To study the effects of genome rearrangements on the organization of bacterial genomes, we used computer simulations that start out with 100 copies of wild type genomes of *E.coli K12*, and 100 copies of the wild type genomes of *Salmonella enterica*. We subjected each genome to 1000 independent stochastic genomic rearrangement events, studying three distinct scenarios of genome rearrangement events: *i)* translocation alone, *ii)* inversion alone, and *iii)* translocation + inversion. In any translocation step, a given number of consecutive metabolic genes (between 2 to 10) is randomly chosen and is translocated to a randomly chosen position in the genome, while in an inversion step, the genes are left in place but the relative ordering

of the genes is reversed. After each rearrangement step, we quantified the robustness of each of the 100 genomes to tandem gene deletions, as determined by the fraction of encoded metabolisms viable on glucose and acetate. Moreover, we identified essential genes in each genome and studied whether these genes are clustered.

In all three scenarios, genome rearrangement events gradually reduce, for both species, robustness to tandem gene deletions on glucose (Figure S22a-c). Inversion alone reduces robustness to a lesser extent than the other two scenarios. Moreover, translocation reduces the clustering of essential genes, but inversion alone does not affect this clustering (Figure S22d-f). The same holds on acetate instead of glucose (Figure S23). Moreover, we observed qualitatively similar patterns when we rearranged larger or smaller numbers of genes in each genome rearrangement event, even though robustness to tandem deletion decreases faster when more genes are rearranged in any one event (figure S24). Further analysis shows that inversion reduces robustness only by shrinking clusters of non-essential genes, whereas translocation both disrupts essential gene clusters and shrinks clusters of non-essential genes (see text S4 and figures S25 and S26).

Text S7: Contribution of operons to the clustering of essential metabolic genes

To examine the contribution of operons in the clustering of essential metabolic genes, we applied partial randomization of metabolic gene orders in the *E.coli K12, Bacillus subtilis* and *Salmonella entrica* genomes. we observed that in a partially reshuffled genome, where only the positions of operonic genes are randomly reshuffled, the clustering of essential genes is as low as in completely reshuffled genomes (Figure S27). In contrast, when we only reshuffled the positions of non-operonic genes, the clustering of essential genes remains similar to that of a wild-type genome. These observations raise the possibility that operons might contribute substantially to the clustering of essentail genes in bacterial genomes.

We then observed that in each of these bacterial genomes, approximately 40% and 20% of operons contain at least one conditionally essential metabolic gene and at least one strictly essential metabolic gene, respectively (table S12). Importantly, we observed that essential metabolic genes are more likely to be part of an operon than other metabolic genes (figures S28a and S29, and tables S13 and S14). We then partitioned each genome's set of strictly essential metabolic genes into operonic genes and non-operonic ones. Using Kuiper's test, we quantified the clustering of genes in each group separately. Operonic essential genes are significantly clustered in all bacterial genomes, but non-operonic genes are not significantly clustered in any genome (Figure S28b and table S15). The same association exists when we consider genes that are essential for viability on a single carbon source such as glucose (figure S30).

How do operons contribute to the clustering of essential genes? As is shown in tables S16 and S17, essential operons themselves are not clustered, but most essential operons contain multiple essential metabolic genes, which contribute to the clustering of essential genes (figure S31). In other words, the fraction of essential operons with multiple essential metabolic genes is considerably higher than expected by chance (figure S31). Preferential preservation of operons with multipe essential genes, despite the fact that operons have continually been destroyed, rebuilt and reshuffled in the course of evolution (15), indicates strong selective advantages associated with including multiple essential genes in the same operon. Thus, selection for co-regulation of multiple essential genes in the same operon has played a substantial role in the non-uniform organization (i.e. clustering) of essential genes in bacterial genomes.

Text S8: Excess robustness to tandem deletion and the organization of metabolic pathways and functional subsystems

Excess robustness to tandem deletion might be exclusively explained by the organization of functionally related genes, for example because genes belonging to the same metabolic pathway tend to be in the same operon (16). To find out whether this is the case, we aimed to quantify the number of metabolic pathways and functional metabolic subsystems that are affected by a given multi-gene deletion. To do so, we pursued two different approaches. In the first, we assigned metabolic genes in a given genome to a metabolic pathway based on the definition of metabolic pathways in the KEGG database (http://www.genome.jp/kegg/pathway.html). In the second, we assigned metabolic genes to metabolic subsystems, which are generalizations of metabolic pathways that are increasingly used in annotating genomes (17). Because metabolic pathways are finer-grained categorizations, the total number of metabolic pathways in a given species is higher than the total number of functional subsystems. For example, while the metabolism of *Escherichia coli K-12 MG1655 (iJO1366)* can be subdivided into 119 different metabolic pathways, it comprises only 42 functional (metabolic) subsystems. With the help of the KEGGREST R package (18) we were able to systematically assign metabolic genes to metabolic pathways for 38 of the 55 genomes in this study. In contrast, metabolic subsystem information is available for all 55 genomes, because for any given metabolic model in the BiGG database, each metabolic reaction has been assigned to at least one functional subsystem (19). These two approaches allow us to easily determine the set of pathways or subsystems that are affected by any given deletion, once we have translated deleted metabolic genes into deleted metabolic reactions.

We first quantified the number of metabolic pathways and functional metabolic subsystems that are affected by a given multi-gene deletion. For this purpose, we compared the average

number of metabolic pathways affected per operon deletion ($\overline{NO}_P$) in each genome with the average number of metabolic pathways affected per tandem deletion ($\overline{NT}_P$) and random deletion ($\overline{NR}_P$) of the same number of genes as the average operon. Specifically, we quantified weighted averages for tandem deletions and random deletions as follows:

$\overline{NT}_P = \sum_{n=2}^{20} w_n \, (NT_P^n)$ and $\overline{NR}_P = \sum_{n=2}^{20} w_n \, (NR_P^n)$,

where w_n is the fraction of operons with *n* metabolic genes in the analyzed genome, and (NT_P^n) and (NR_P^n) are the average numbers of metabolic pathways affected by tandem and random deletion of *n* metabolic genes, respectively. Note that we allow *n* to vary between 2 and 20, because that is the range of gene numbers in the smallest and largest operons in our study genomes.

Similarly, we compared the average number of subsystems affected per operon deletion ($\overline{NO}_{sub}$) in each genome with the average number of subsystems affected per tandem deletion ($\overline{NT}_{sub}$) and random deletion ($\overline{NR}_{sub}$) of the same number of genes as the average operon. Again, we quantified weighted averages for tandem deletions and random deletions as follows:

$\overline{NT}_{sub} = \sum_{n=2}^{20} w_n \, (NT_{sub}^n)$ and $\overline{NR}_{sub} = \sum_{n=2}^{20} w_n \, (NR_{sub}^n)$,

where w_n is the fraction of operons with *n* metabolic genes in the analyzed genome, and (NT_{sub}^n) and (NR_{sub}^n) are the average numbers of subsystems affected by tandem and random deletion of *n* metabolic genes, respectively.

As expected, we observed that metabolic genes belonging to the same operon are functionally more related and their simultaneous deletion thus affects fewer metabolic pathways or subsystems than tandem and random deletions of the same length (figure S32). Thus, the observation that tandem deletions of *n* genes consistently affect fewer metabolic subsystems than random deletions of the same *n* genes (figures S33 and S34) is at least partly a consequence of the organization of metabolic genes in operons.

Next, to examine the effect of the number of affected pathways or subsystems per deletion, we partially randomized genomes in two different ways, which allowed us to study two different types of deletions. In partially randomized genomes of type I, we kept the strictly essential genes in the same position as in a wild-type genome, but we randomly reshuffled all other genes. This type I randomization preserves the clustering of essential genes but allows the organization of metabolic pathways and subsystems to become randomized. We then determined the number of metabolic pathways and subsystems affected by tandem deletions of a given number of genes, referring to such deletions as partially random deletions of type I.

We observed that the number of affected metabolic pathways and subsystems is considerably higher than that affected by tandem deletion in a non-shuffled (wild-type) genome (Figures S35 and S36). In contrast, in partially randomized genomes of type II, we kept the organization of metabolic pathways or subsystems intact, while allowing the position of essential metabolic genes to be reshuffled with non-essential metabolic genes. We performed this analysis twice: First, we kept the organization of "metabolic pathways" intact, and second, we kept the organization of "metabolic subsystems" intact. In the analysis, we reshuffled for each metabolic pathway (or subsystem) the position of the subset of (non-shuffled) metabolic genes that belong to the given pathway (or subsystem). By repeating this procedure for all metabolic pathways (or subsystems), we ensured that the position of all genes (including essential ones) could get randomized while leaving the genomic organization of metabolic pathways (or subsystems) intact. We subjected type II randomized genomes to tandem gene deletions, referring to such deletions as type II partially random deletions.

We then examined the robustness of *E.coli* metabolism to both type I and type II partially random deletions, which allowed us to de-convolve the contribution of the organization of metabolic pathways (or subsystems) from that of the organization of essential genes to the excess robustness to tandem deletions. Figures 3a (in main text), S37 and S38, show that robustness to type I deletions is consistently higher than robustness to type II deletions. While the robustness to partially random deletion of type I is similar to that of robustness to tandem deletions (in a non-shuffled genome), the robustness to partially random deletion of type II is closer to robustness to fully random deletion (Figures 3a (in main text), S37 and S38).

To make this comparison more quantitative, we determined which fraction of the excess robustness to tandem deletion is preserved ($F(n)$) in each type of partially random deletion as follows:

$$F(n) = \frac{R_{partial}(n) - R_{random}(n)}{R_{tandem}(n) - R_{random}(n)}$$

Here, $R_{tandem}(n)$, $R_{random}(n)$ and $R_{partial}(n)$ refer to robustness to tandem, random and partially random deletion of n metabolic genes, respectively. We observed that a consistently higher fraction of the excess robustness to tandem deletion is preserved in partially random deletions of type I compared to type II (Figures 3b (in main text), S39 and S40). This indicates that the organization of essential genes is more important for robustness to tandem deletion, whereas the organization of metabolic pathways (or subsystems) exerts a smaller effect. In conclusion, although the effect of the genomic arrangement of metabolic pathways (or subsystems) is not negligible, its impact on the excess robustness to tandem deletion is less important than the organization of essential genes in bacterial genomes.

Note that in our partially randomized deletion of both types, we have performed tandem deletion on the partially randomized genome and referred to it as "partially random deletion". We can also perform random deletion on these partially randomized genomes. However, since genes to be deleted in random deletions are selected regardless of their genomic positions, it is not surprising to see that robustness to random deletion in partially randomized genomes is close to robustness to random deletion in the wild type genome.

We subjected the partially randomized genomes of both types to both tandem and random deletions. Figure S41 indicates that whereas in partially randomized genomes of type I (upper panels), the effects of tandem and random deletion differ substantially, in partially randomized deletions of type II (lower panels), this difference is much reduced. This observation further underscores the greater importance of essential gene organization for robustness to multi-gene deletions as compared to the importance of the genomic organization of metabolic pathways.

Text S9: Repulsion of synthetic lethal gene pairs is not simply due to the genomic repulsion of pairs of metabolic subsystems or pathways

We wanted to find out whether the repulsion of synthetic lethal gene pairs may simply be a by-product of the organization of metabolic subsystems or pathways. This could occur if genes in synthetic lethal pairs may preferentially occur in different metabolic subsystems or pathways, and genes encoding these subsystems or pathways may be located far from each other in the genome.

To examine this possibility, we considered all distinct pairs of subsystems for each of the 55 genomes in this study. For example, in *Escherichia coli K-12 MG1655 (iJO1366)*, there are 42 distinct metabolic subsystems, so we considered all $\binom{42}{2} = 984$ distinct pairs of subsystems. Then, for each given pair of subsystems, we determined the minimum distance (i.e. D_{min}; in terms of the number of intervening metabolic genes) between the two sets of genes encoding the two metabolic subsystems. In *Escherichia coli K-12 MG1655 (iJO1366)*, the majority of the subsystem pairs are separated by a short genomic distance. For example the median of D_{min} over the 984 pairs of metabolic subsystems is 2, and the D_{min} of 93% of the subsystem pairs is smaller than 50. Only 69 subsystem pairs (7%) were in repulsion ($D_{min} > 50$). Similar observations hold for other genomes (table S18). Thus, the majority of the subsystem pairs is not in repulsion but is located close to each other.

We repeated this analysis based on metabolic pathways, and did so for those 38 genomes whose pathway information is available in the KEGG database. For each genome, we again considered all possible pairs of metabolic pathways, and for each pair we determined the

minimum distance (i.e. D_{min}) between the two sets of genes encoding the two metabolic pathways. We observed that in all genomes, the majority of the pathway pairs are separated by a short genomic distance (table S19). For example, in *Escherichia coli K-12 MG1655 (iJO1366)*, the median of D_{min} over the $\binom{119}{2} = 7021$ pairs of metabolic pathways is 4, and the $D_{min}$of 88% of the pathway pairs is smaller than 50. Only 842 pathway pairs (12%) are in repulsion ($D_{min} > 50$). Thus, the majority of pathway pairs are not in repulsion but are located close to each other.

We then examined the possibility that the synthetic lethal gene pairs might belong to the rare pairs of subsystems or pathways that are in repulsion (with $D_{min} > 50$). We observed, first, that for the majority of the synthetic lethal gene pairs, both genes belong in fact to the same metabolic subsystem or pathway. For example, in *Escherichia coli K-12 MG1655 (iJO1366)*, both genes in 40 out of the 70 synthetic lethal gene pairs (57.14%) belong to the same metabolic subsystem, and both genes in 44 out of the 70 synthetic lethal gene pairs (62.85%) belong to the same metabolic pathway. Repulsion between metabolic subsystems or pathways could not possibly explain the repulsion of these synthetic lethal pairs. Secondly, for each of the 30 remaining synthetic lethal gene pairs that are not in the same subsystem, $D_{min} \leq 5$ (table S18), and for the 26 remaining synthetic lethal gene pairs that are not in the same pathway, $D_{min} = 1$ (table S19). That is, none of these subsystems are in repulsion. Similar observations hold for the other genomes (tables S18 and S19). In sum, the repulsion of synthetic lethal gene pairs does not simply result from metabolic pathway or subsystem organization.

Text S10: The uber-operon hypothesis and the excess robustness to tandem deletion

Another candidate reason for an excess robustness to tandem deletion is that operons may form higher order functional units, such as uber-operons, in which a set of functionally related or co-regulated clustered together in the genome. To find pertinent evidence, we first asked whether essential operons are themselves clustered in the genome or not. We observed that neither conditionally essential operons (i.e. operons whose deletion abolishes viability on at least one carbon source) nor strictly essential operons (i.e. operons whose deletion abolishes viability on all the carbon sources on which the corresponding wild-type genome is viable) are clustered in the genome (tables S16 and S17). However, if uber-operons exist and strongly affect robustness to tandem deletion, we would expect that deletion of two or more nearby (functionally related) operons is less deleterious than the deletion of the same number of operons randomly located in the genome. To validate this hypothesis, we generated deletional variants in which the units of deletion were operons instead of genes, and compared

robustness to the tandem deletion of a given number of operons with robustness to the random deletion of the same number of operons. Specifically, for each of three genomes, we systematically generated deletional variants in which n operons ($1 \leq n \leq 10$) are deleted. For each n, we compared (i) robustness to tandem operon deletions, where n consecutive operons are deleted, with (ii) robustness to random operon deletions, where n randomly chosen operons (without regard to their positions) are deleted. We quantified conditional and strict robustness as the fraction of deletional variants retaining viability on at least one carbon source and on all carbon sources, respectively. Figures S42 and S43 show that robustness to tandem operon deletion and to random operon deletion differ only slightly from each other, indicating that uber-operons are not important contributors to the excess of robustness to tandem deletions.

Finally, we also quantified the importance of the relative ordering of operons, as suggested by the uber-operon hypothesis on the clustering of essential genes. In a partial reshuffling analysis of three different genomes, we reshuffled the ordering of operons without changing the orders of genes inside any given operon. This reshuffling reduced the clustering of essential metabolic genes compared to wild-type genomes (Figure S44). However, this clustering was still considerably higher than in completely randomized genomes (Figure S44). These observations suggest that the relative arrangement of operons in the genome, and a higher order organization of operons, as suggested by the uber-operon hypothesis (20), influence the clustering of essential metabolic genes to some extent.

Text S11: The selfish operon hypothesis and the excess robustness to tandem deletion

Another potential candidate to explain the excess robustness to tandem deletion is the selfish operon hypothesis (21), which was proposed to explain the clustering of functionally related genes into operons. This hypothesis asserts that the organization of genes into operons is not necessarily beneficial for a host genome, but for the constituent genes, because an operon enables the spreading of its genes to new cells and species by horizontal gene transfer. The hypothesis, which has been criticized before (22, 23), also predicts that horizontally transferred operons would harbor genes with peripheral (i.e. non-essential) metabolic functions (21), such that their deletion would be more tolerable than the deletion of other genes in the genome.

To validate this prediction, we identified all operons and their corresponding metabolic genes in 52 bacterial genomes using the DOOR database(24). For each genome, we generated all operon deletion variants, i.e., we selected one of the operons and deleted all metabolic genes belonging to it, and repeated this procedure for all operons. We then used flux balance analysis to determine the viability of each operon deletion variant on 102 distinct carbon sources. We

quantified the robustness of a given genome to operon deletion ($R_{Operonic}$) as the fraction of operon deletion variants that retain viability in a given environment or environments. To compare this robustness with the average robustness to tandem and random deletions of the same length for a given bacterial genome, we quantified the weighted average of robustness to tandem ($\bar{R}_{tandem}$) and random ($\bar{R}_{random}$) deletion which we define respectively as $\bar{R}_{tandem} = \sum_{n=2}^{20} w_n R^n_{tandem}$, and $\bar{R}_{random} = \sum_{n=2}^{20} w_n R^n_{random}$, where w_n is the fraction of operons with *n* metabolic genes in the analyzed genome, and R^n_{tandem} and R^n_{random} are the genome's robustness to tandem and random deletions of *n* metabolic genes. Note that *n* varies between 2 and 20, because the smallest and largest operons in our study genomes have this respective number of genes. This weighted average ensures that the average length of tandem (or random) deletions that enter the calculation is the same as the average length of operons in any given genome. Figure S45 and table S21 show that robustness to operon deletion is slightly higher than the average robustness to tandem deletion, but the difference is not as dramatic as that between robustness to tandem deletion and random deletion. Thus, the dispensability of operons as implied by the selfish operon theory cannot fully explain the excess robustness to tandem deletions.

Text S12: Flux balance Analysis

Flux balance analysis (FBA) is a widely used computational method for the quantitative analysis and modeling of metabolic networks (25). FBA predicts the metabolic flux through each reaction in a given metabolic network using the stoichiometric coefficients of metabolites participating in the network's reactions. Stoichiometric coefficients are stored in a stoichiometric matrix *S*, which is of dimension $m \times n$, where *m* and *n*, respectively, denote the number of metabolites and the number of reactions in a metabolic network. FBA constrains the flux through each reaction based on the assumption that the metabolic network is in a steady state in which metabolite concentrations do not change, i.e., $Sv = 0$, where *v* is the vector of metabolic fluxes v_i through reaction *i*. The solutions of the equation $Sv = 0$, that is, the null space of the matrix S, comprises all flux vectors that are allowable in steady state. The null space can be further constrained by physicochemical information regarding the maximally and minimally possible fluxes through each reaction. FBA relies on linear programming to identify those allowable flux vector(s) that maximize an objective function *Z*. This task can be formulated as finding a flux vector v^* with the property

$v^* = max_v\, Z(v) = max_v \{ c^T v \mid Sv = 0,\ a \leq v \leq b \}$,

where the vector *c* contains a set of scalar coefficients representing the maximization criterion, and each entry a_i and b_i of vectors a and b, respectively, indicate the minimally and maximally possible flux through reaction *i*. The vector *c* represents the proportions of each

small biomass molecule in a cell's biomass. The quantity v^* maximizes the biomass growth flux, that is, the rate at which a metabolic network can produce biomass (25). Here we use FBA to predict qualitatively whether a given metabolic network is viable in a given environment, and we consider a metabolic network viable if it can produce all essential biomass precursors. In a free-living bacterium like *E.coli,* there are approximately 60 such molecules including 20 amino acids, DNA, and RNA precursors, lipids and cofactors. We used the biomass composition of the *E. coli* metabolic model iAF1260 to define the vector c (26). Moreover, we used the packages CPLEX (11.0, ILOG; http://www.ilog.com/) and CLP (1.4, Coin-OR; https://projects/coin-or.org/Clp) to solve the linear programming problem of FBA. The major limitation of FBA is that it neglects regulatory constraints that can arise through suboptimal expression or regulation of enzymes. Newly horizontally transferred genes cannot easily establish regulatory interactions with their host genes, and it may thus take considerable adaptive evolution until they become expressed at a maximal or optimal level (27). Such regulatory constraints would be especially important if we focused on quantitative predictions of biomass growth (28). However, we use FBA solely for qualitative predictions of viability. This focus on qualitative phenotypes is biologically sensible. The reason is that many organisms grow slowly in their native environment (29), implying that regulation for maximal biomass production is far from universal. Moreover, we note that regulatory constraints can easily be broken in evolution, even on the short time scales of laboratory evolution experiments (28, 30, 31).

Text S13: Minimal environments used in this study

Among the 55 bacterial metabolisms we study, we identified 137 unique carbon-containing metabolites that occur in the metabolism of all species. Thus, we considered 137 minimal growth environments that were distinguished by these carbon sources. Each of these environments included one carbon source, as well as oxygen, ammonium, inorganic phosphate, sulfate, sodium, potassium, cobalt, iron (Fe2+ and Fe3+), protons, water, molybdate, copper, calcium, chloride, magnesium, manganese, and zinc. In other words, we varied the carbon source while keeping all other nutrients constant. None of the 55 species we studied were viable on 34 of these 137 minimal environments, so we excluded the corresponding carbon sources from this study, and performed our analysis with the remaining 103 minimal environments, whose carbon sources are listed in table S2.

Supplementary Figures

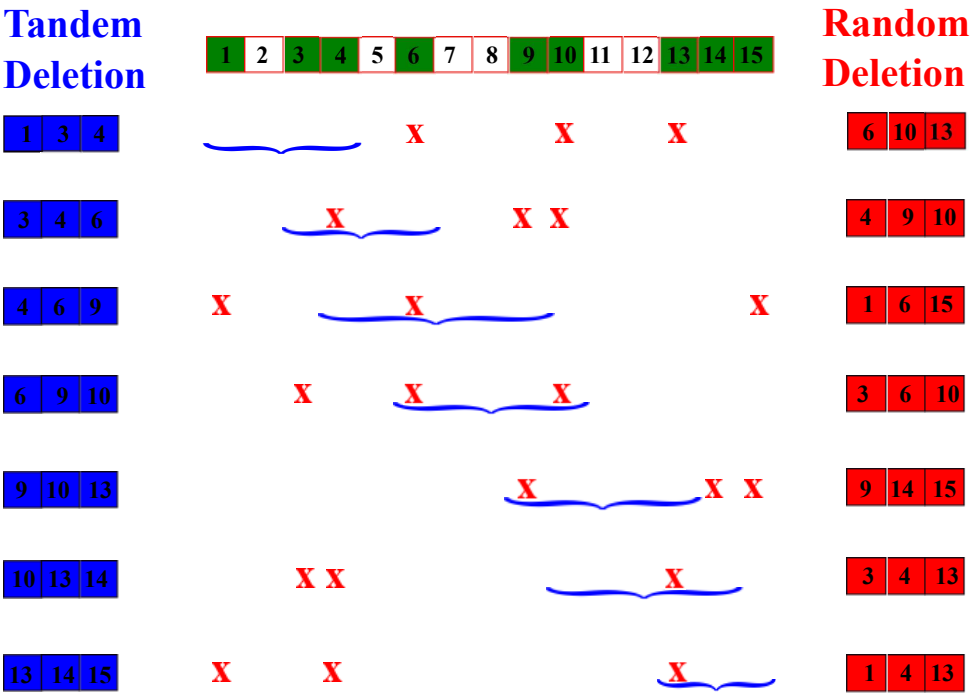

Figure S1: Tandem deletion versus random deletion of metabolic genes. A hypothetical linear genome with 15 genes, 9 of which are metabolic genes (green). To quantify deletional robustness to tandem deletions of 3 genes, we consider all contiguous segments of the genome that contain 3 metabolic genes (blue curly brackets). This results in 7 distinct deletional variants with three deleted metabolic genes (blue boxes on the left). To quantify random deletion, deletional variants (red boxes on the right) are generated by deleting 3 genes that are randomly chosen among all metabolic genes with uniform probability (red crosses). Since our computational approach only pertains to metabolic genes, we do not delete non-metabolic genes. In other words, all deletional variants retain all non-metabolic genes. Note that bacterial genomes are circular, so in a circular version of this example, we would need to consider two additional deletional variants, i.e. tandem deletions {14,15,1} and {15,1,3} would also have to be generated, and two additional random deletional variants would also have to be considered.

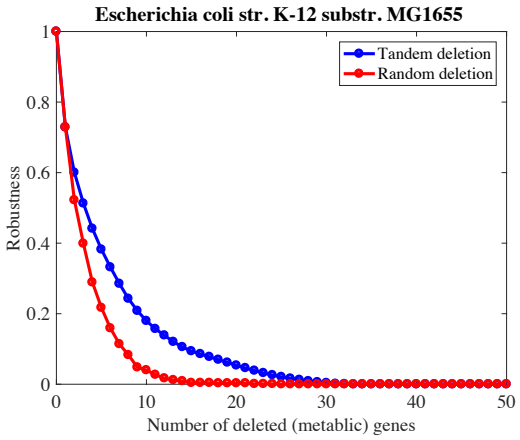

Figure S2: Robustness to tandem deletion versus random deletion (strict phenotype definition). The vertical axis shows the robustness of *Escherichia coli K-12 MG1655 (iJO1366)* to tandem (blue) and random (red) deletion of metabolic genes, averaged over all deletional variants we examined, as a function of the number of deleted genes (horizontal axis). In this analysis, robustness is defined as the fraction of deletional variants that retain viability on all 97 carbon sources on which the wild type *Escherichia coli K-12 MG1655 (iJO1366)* is viable. Interpolation between data points is linear and is displayed as a visual guide

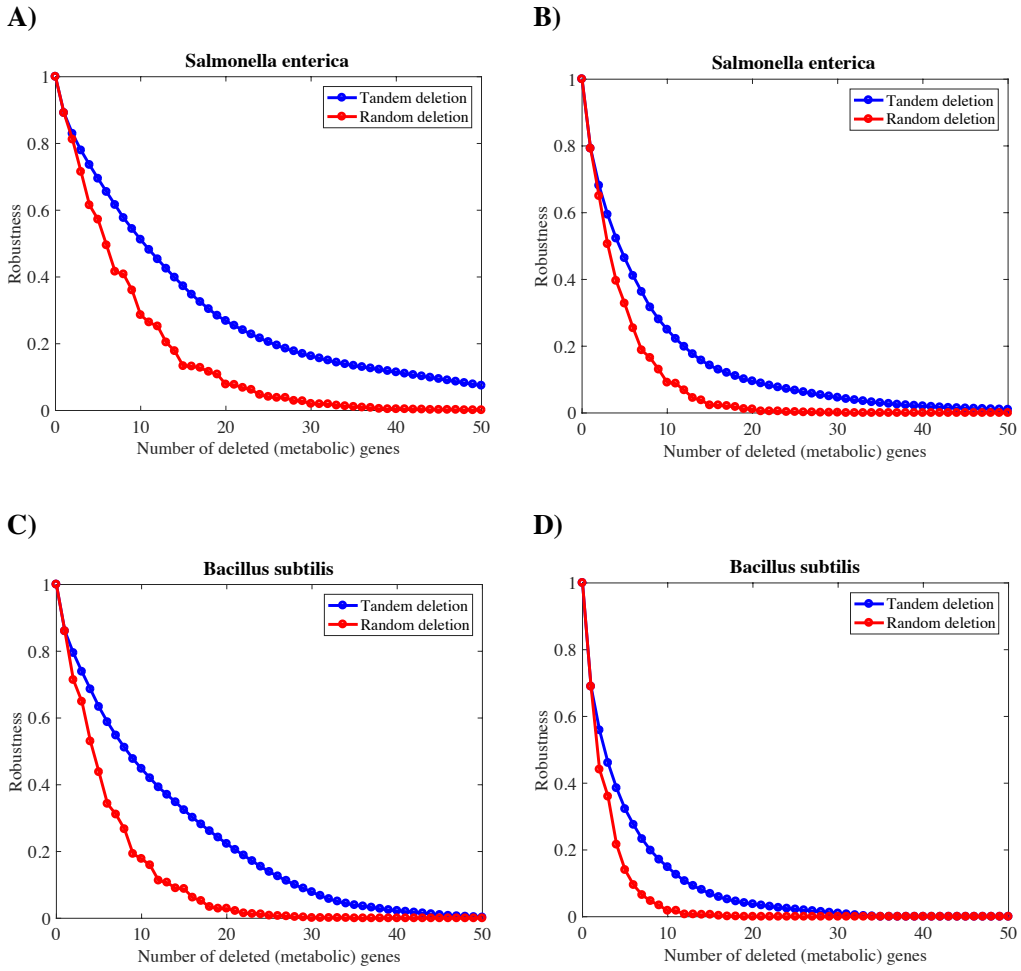

Figure S3: Robustness to tandem deletion versus random deletion (using alternative bacterial genomes). The vertical axes show the robustness of *Salmonella enterica* (panels **A** and **B**) and *Bacillus subtillis* (panels **C** and **D**) to tandem (blue) and random (red) deletion of metabolic genes, averaged over all deletional variants we examined, as a function of the number of deleted genes (horizontal axis). Robustness in panels A and C is defined conditionally, i.e., as the fraction of deletional variants that retain viability on at least one carbon source, while in panels B and D it is defined strictly, i.e., as the fraction of deletional variants that retain viability on all carbon sources on which the wild type metabolism is viable. Interpolation between data points is linear and is displayed as a visual guide.

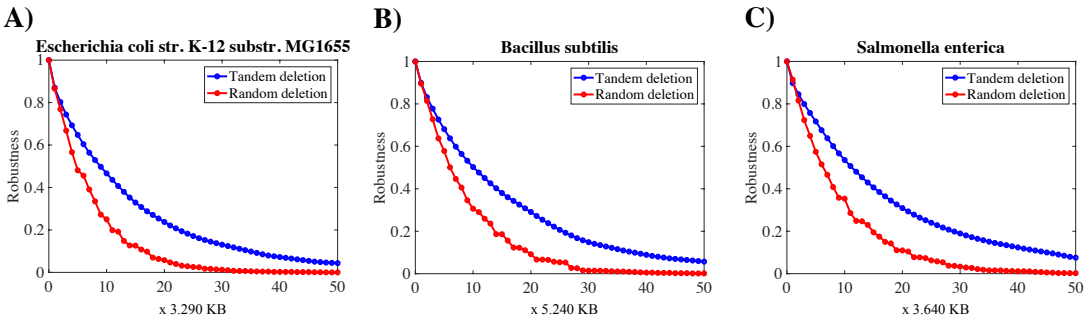

Figure S4: Robustness to tandem deletion versus random deletion (using an alternative deletion approach (see text S2) and conditional definition of robustness). The vertical axes show the robustness of *Escherichia coli K-12 MG1655 (iJO1366)* (panel **A**), *Bacillus subtillis* (panel **B**) and *Salmonella enterica* (panel **C**) to tandem (blue) and random (red) deletion of metabolic genes, averaged over all deletional variants we examined, as a function of the length of deleted segments in kbp (horizontal axis). Note that length units on the horizontal axes differ among the panels, because metabolic genes occupy different amounts of genomic DNA in the three organisms. They are 3.290, 5.240 and 3.640 kilobases, respectively, for panels A to C. Length units are chosen such that each length unit contains, on average, one metabolic gene in each of the genomes. The length of deleted DNA is expressed as a multiple of this length unit. In this analysis, robustness is defined conditionally, i.e., as the fraction of deletional variants that retain viability on at least one carbon source. Interpolation between data points is linear and is displayed as a visual guide.

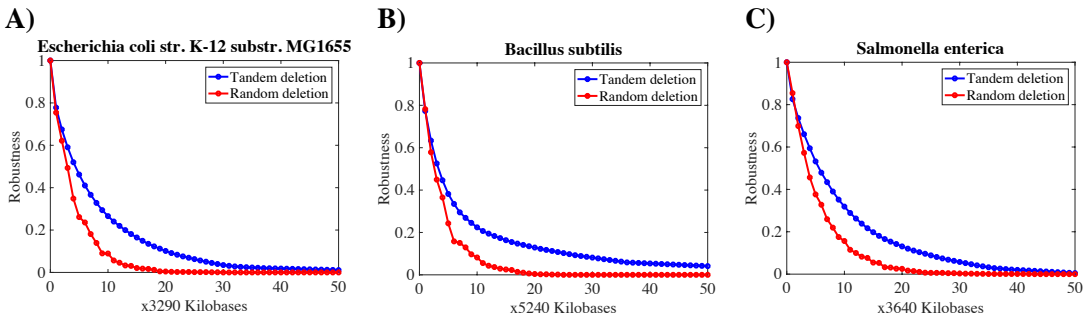

Figure S5: Robustness to tandem deletion versus random deletion (using an alternative deletion approach (see text S2) and strict definition of robustness). The vertical axes show the robustness of *Escherichia coli K-12 MG1655 (iJO1366)* (panel **A**), *Bacillus subtillis* (panel **B**) and *Salmonella enterica* (panel **C**) to tandem (blue) and random (red) deletion of metabolic genes, averaged over all deletional variants we examined, as a function of the length of deleted segments in kbp (horizontal axis). Note that length units on the horizontal axes differs among the panels, because metabolic genes occupy different amounts of genomic DNA in the three organisms. They are 3.290, 5.240 and 3.640 kilobases, respectively, for panels A to C. Length units are chosen such that each length unit contains, on average, one metabolic gene in each of the genomes. The length of deleted DNA is expressed as a multiple of this length unit. In this analysis robustness is defined strictly, i.e., as the fraction of deletional variants that retain viability on all carbon sources on which the wild type metabolism is viable. Interpolation between data points is linear and is displayed as a visual guide.

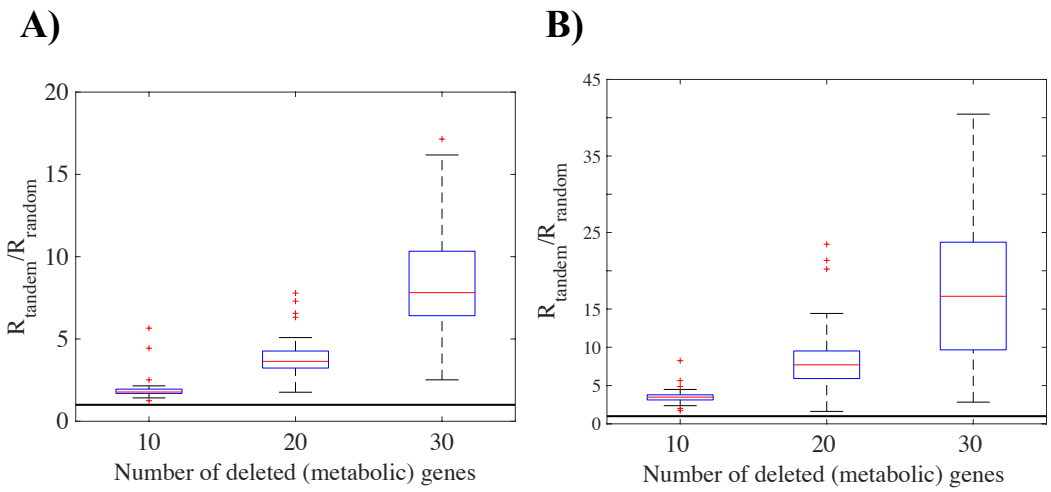

Figure S6: Excess robustness to tandem deletion. The vertical axes show the excess robustness to tandem deletion, defined as the ratio of robustness to tandem deletion and robustness to random deletion (R_{tandem}/R_{random}), for all 55 bacterial genomes, as a function of the number of deleted genes (horizontal axis). In panel A robustness is defined as the fraction of deletional variants that retain viability on all carbon sources on which the wild type metabolism is viable. Paired-sample t-tests between *i)* n=10 and n=20 yield P $<10^{-12}$, *ii)* between n=10, n=30 yield P $<10^{-13}$, and *iii)* between n=15, n=20 yield P$<10^{-7}$. In panel B robustness is defined as the fraction of deletional variants that retain viability on at least one carbon source. Paired-sample t-tests between *i)* n=10 and n=20 yield P$<10^{-23}$, *ii)* between n=10, n=30 yield P$<10^{-7}$, and *iii)* between n=20, n=30 yield P$<10^{-4}$.Boxes span the 25-th to 75-th percentile, whiskers indicate maxima and minima, and red '+' signs show outliers.

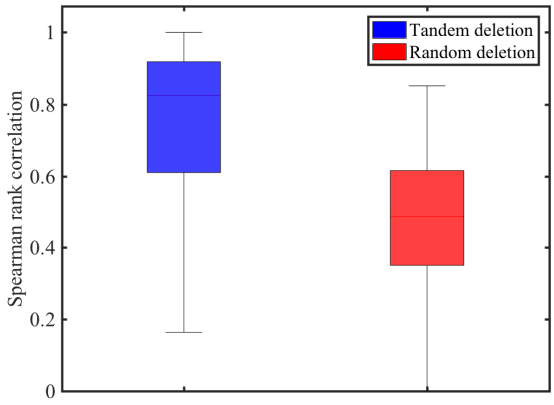

Figure S7: Higher correlation among carbon sources for tandem robustness as compared to random robustness. The same metabolism may show different robustness to tandem or random deletions of a given number of genes, depending on the carbon source environment in which this robustness is evaluated. We computed the robustness (R_{tandem}) of all our 55 prokaryotic metabolisms to tandem deletions of five genes on all carbon sources on which these metabolisms are viable, and then determined Spearman's rank correlation coefficients between R_{tandem} on these carbon sources, for all pairs of metabolisms. We performed an analogous calculation for robustness (R_{random}) to random deletions of five genes. Boxes indicate the distribution of Spearman's rank correlation coefficient of robustness to tandem deletions of five genes (R_{tandem}, blue) and random deletions of five genes (R_{random}, red) for those carbon sources on which both metabolisms in a given pair are viable. Paired-sample t-test between Spearman rank correlation coefficients of tandem versus random deletion yields P$<10^{-15}$. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. In box-whisker plots, boxes span the 25-th to 75-th percentile, whiskers indicate maxima and minima, and red '+' signs show outliers.

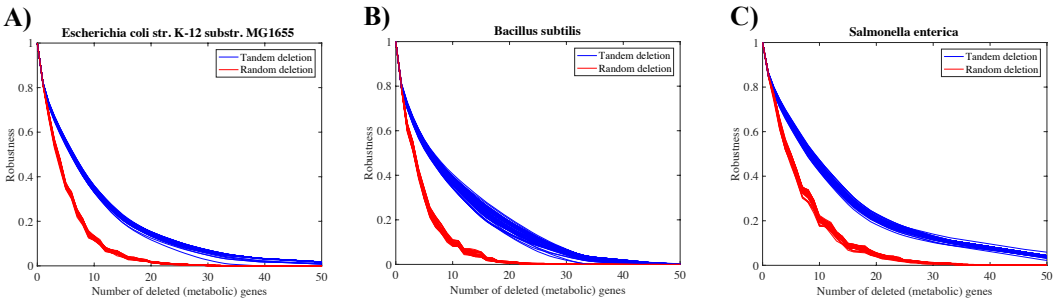

Figure S8: Higher variability among carbon sources in robustness to tandem deletion than random deletion. Each panel shows **A)** 97 blue curves and 97 red curves, **B)** 68 blue curves and 68 red curves, and **C)** 87 blue curves and 87 red curves indicating robustness to tandem (blue) and random (red) deletion for each of the A) 97, B) 68, and C) 87 carbon sources on which A) *Escherichia coli K-12 MG1655 (iJO1366)*, B) *Bacillus subtillis* and C) *Salmonella enterica* are viable as a function of the number of deleted metabolic genes (horizontal axis). Each curve is obtained by a linear interploation betrween 50 data points.

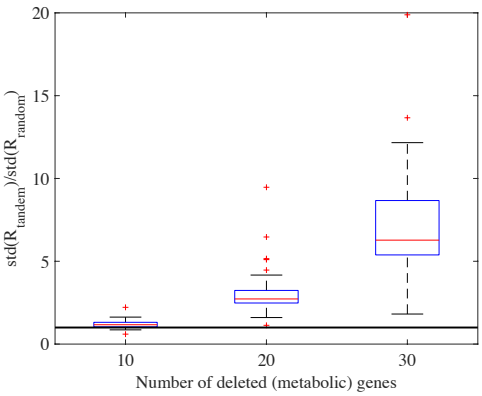

Figure S9: Excess variability in robustness to tandem deletions. The vertical axis shows the excess variability in robustness to tandem deletion among different carbon sources defined as the ratio of the standard deviation of robustness to tandem deletion (among different carbon sources) and the standard deviation of robustness to random deletion ($Std(R_{tandem})/Std(R_{random})$), for all 55 bacterial genomes, using three different number of deleted genes (horizontal axis). Boxes span the 25-th to 75-th percentile, whiskers indicate maxima and minima, and red '+' signs show outliers. Paired-sample t-test between *i)* n=10 and n=20 has P-value<10^{-16}, *ii)* n=10, n=30 has P-value<10^{-12}, and *iii)* n=20, n=30 has P-value<10^{-10}).

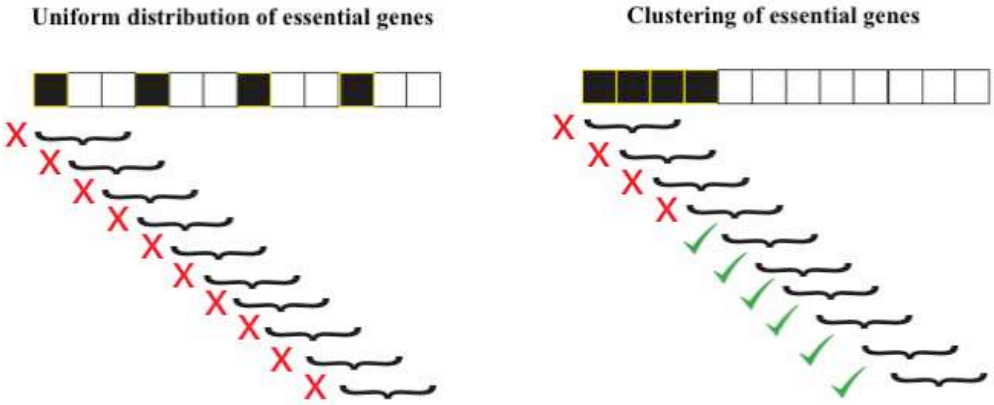

Figure S10: Clustering of essential genes can enhance robustness to tandem gene deletion. The figure illustrates two different hypothetical genome organizations, each of which has the same number of genes (12) and the same number of essential genes (4, shown as black boxes). Whereas in the left genome the essential genes are uniformly distributed, in the right genome they are clustered (i.e. conentrated in one region of the genome). Each curly bracket indicates tandem deletions of 3 specific consecutive genes. Red crosses and green check marks indicate whether any one deletion would disrupt or preserve viability. In the left genome, each tandem deletion includes one essential gene and will thus disrupt viability. In contrast, for the right genome, only the left-most four deletional variants include an essential gene, implying that 60% of deletions preserve viability, such that the robustness to deletions

is 0.6. Note that we have made the simplifying assumption that simultaneous deletion of multiple non-essential genes is not lethal.

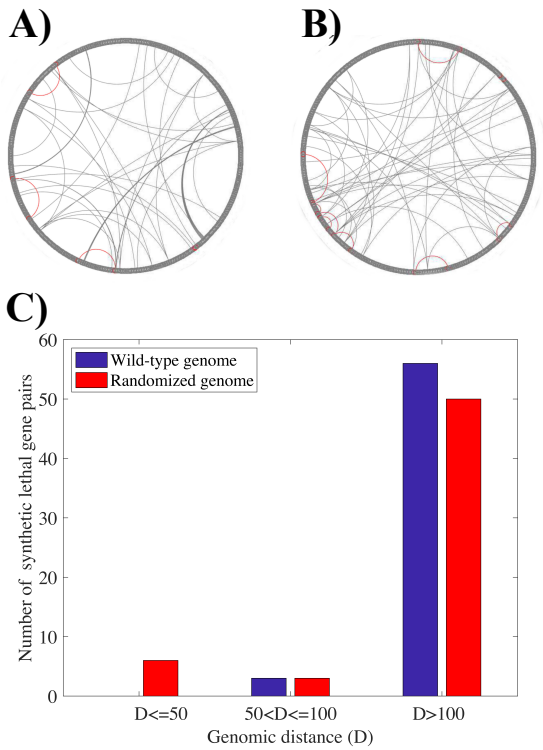

Figure S11: Repulsion of synthetic lethal genes in the *E.coli 083:H1* genome. A) Circos plot of *Escherichia coli K-12 083:H1* genome, in which metabolic genes are arranged according to their order in the genome. An arc connects two genes if they form an unconditionally synthetic lethal pair. An arc is colored red if the genomic distance (in number of intervening genes) between two synthetic lethal gene pairs is less than 100. **B)** Same as A, but for randomized gene order. Note the many short-range synthetic lethality interactions after gene order randomization. **C)** Barplot of the genomic distance (in number of intervening genes) between unconditionally synthetic lethal metabolic gene pairs in the wild-type (blue) and randomized (yellow) *Escherichia coli K-12 083:H1* genome. Note the lack of short-distance synthetic lethal pairs with fewer than 50 intervening genes in the wild type genome (Fisher's exact-test: P=0.0220 and adjusted P=0.0306; See also table S7).

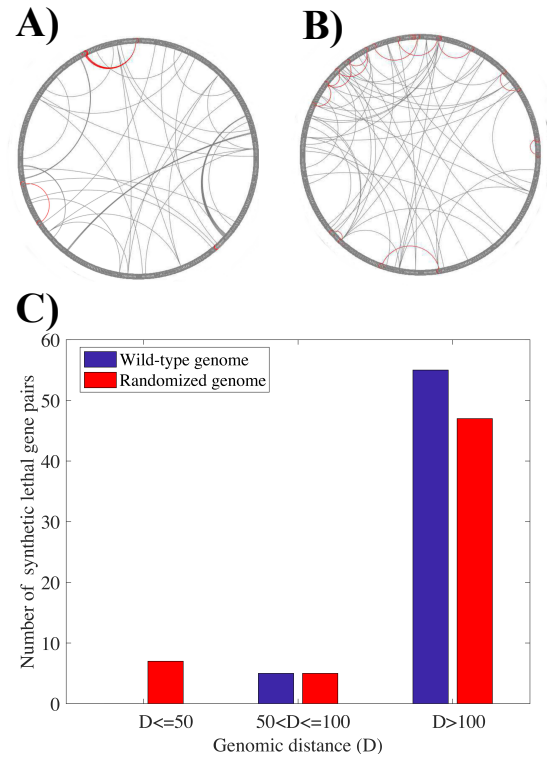

Figure S12: Repulsion of synthetic lethal genes in the *Shigella flexneri* genome. A) Circos plot of *Shigella flexneri* genome, in which metabolic genes are arranged according to their order in the genome. An arc connects two genes if they form an unconditionally synthetic lethal pair. An arc is colored red if the genomic distance (in number of intervening genes) between two synthetic lethal gene pairs is less than 100. **B)** Same as A, but for randomized gene order. Note the many short-ranged synthetic lethality interactions after gene order randomization. **C)** Barplot of the genomic distance (in number of intervening genes) between unconditionally synthetic lethal metabolic gene pairs in the wild-type (blue) and randomized (yellow) *Shigella flexneri* genome. Note the lack of short-distance synthetic lethal pairs with fewer than 50 intervening genes in the wild type genome. (Fisher exact-test: P=0.0138 and adjusted P=0.0306; See also table S7)

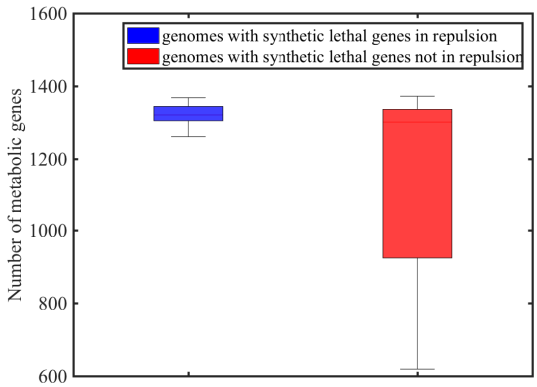

Figure S13: Repulsion of synthetic lethal genes and genome length. Metabolic genomes in which strictly synthetic lethal genes are in repulsion tend to be larger than those in which strictly synthetic lethal genes are not in repulsion (t-test P-value<10^{-4}). In none of our genomes with fewer than 1200 metabolic genes were synthetically lethal genes in repulsion.

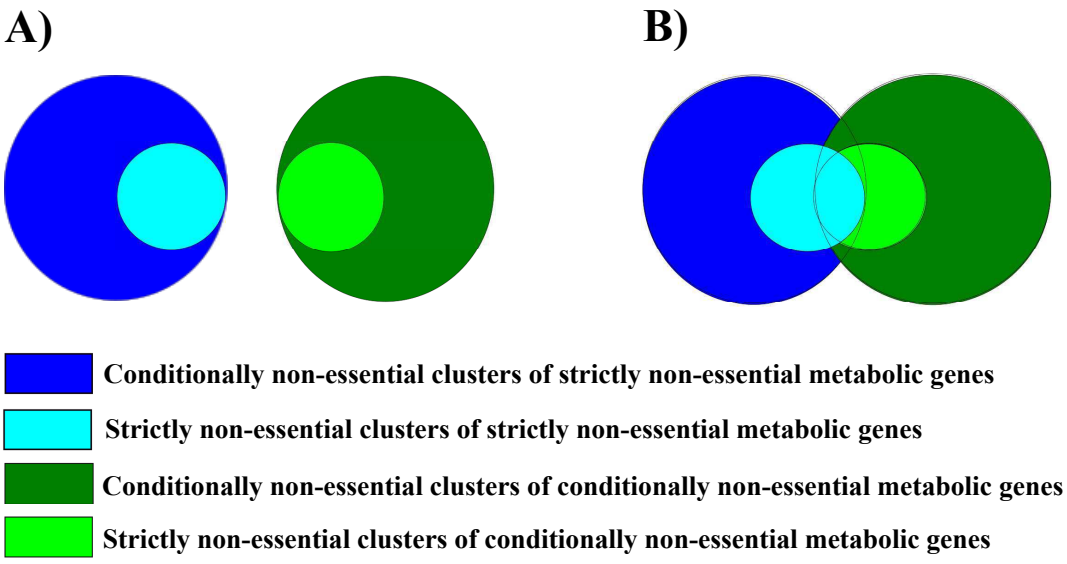

Figure S14: Different types of non-essential gene clusters. Successive strictly non-essential genes might form a cluster of strictly non-essential genes, whose simultaneous deletion does not abolish viability on any carbon source. We refer to such a cluster as a strictly non-essential cluster of strictly non-essential genes (cyan). Alternatively, genes in a strictly non-essential cluster might not abolish viability on at least one carbon source, in which case we would say they form a conditionally non-essential clusters of strictly non-essential genes (blue). The former kind of cluster would always be a subset of the latter kind of cluster, if one views a strictly non-essential gene as a special case of a conditionally non-essential gene.

Likewise, successive conditionally non-essential genes might form a cluster of conditionally non-essential genes, whose simultaneous deletion does not abolish viability on any carbon source. We refer to such a cluster as strictly non-essential cluster of conditionally non-essential genes (light green). Alternatively, genes in a conditionally non-essential cluster might not abolish viability on at least one carbon source, in which case we would say they form a conditionally non-essential cluster of conditionally non-essential genes (dark green). The former is always a subset of the latter.

A cluster of strictly non-essential genes must be flanked by two conditionally or strictly essential genes, whereas a cluster of conditionally non-essential genes must be flanked by two strictly essential genes. If a genome contains no cluster of strictly non-essential genes flanked by two strictly essential genes, the two sets clusters are non-overlapping (panel **A**), but if there is at least one cluster of strictly non-essential genes that are flanked by two strictly essential genes, the two sets overlap (panel **B**).

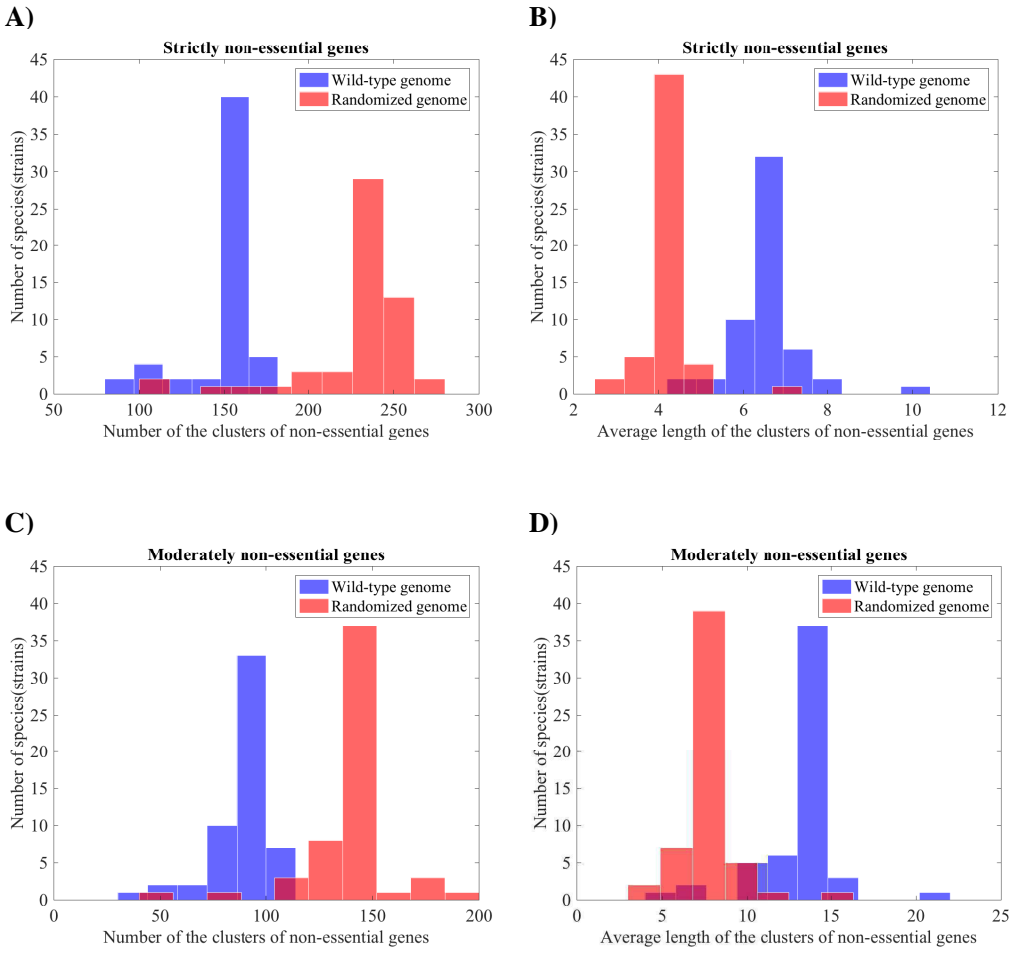

Figure S15: Long clusters of non-essential genes in bacterial genomes. Histogram of the number of clusters of **A)** strictly non-essential metabolic genes, and **C)** conditionally non-essential metabolic genes, and the average length of clusters of **B)** strictly non-essential metabolic genes and **D)** conditionally non–essential metabolic genes, among the 55 wild-type bacterial genomes (blue) and the corresponding 55 randomized genomes (red). We consider a metabolic gene as conditionally non-essential if its deletion does not abolish viability on at least one carbon source, and consider a metabolic gene as strictly non-essential if its deletion does not abolish viability on any carbon source. A *cluster of conditionally non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that intervene between two successive but non-adjacent strictly essential metabolic genes. Likewise, a *cluster of strictly non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that intervene between two successive but non-adjacent conditionally essential metabolic genes.

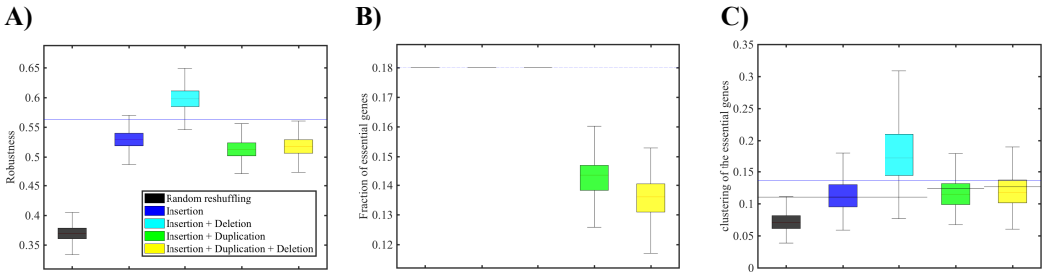

Figure S16: Emergence of essential gene clusters through gradual genomic expansion from a minimal towards a full-sized genome. Data is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Escherichia coli K-12 MG1655 (iJO1366)*, towards a final genome of equal size as the wild-type *Escherichia coli K-12* genome. We simulated genome "growth" in four different ways (see legend): insertion of genes alone (blue), insertion + deletion (cyan), insertion + duplication (green), and insertion + duplication + deletion (yellow). As a control we generated 100 genomes obtained by random reshuffling of the wild-type *Escherichia coli K-12* genome (black). Vertical axes indicate **A)** robustness to tandem deletions of five genes, **B)** fractions of essential genes, and **C)** clustering of essential genes in the final 100 full-sized genomes, as indicated by Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line in panel A) indicates metabolic robustness of the wild type *Escherichia coli K-12* genome to tandem deletions of five metabolic genes in glucose minimal medium. The blue horizontal line in C) shows the clustering of essential metabolic genes in the wild type *Escherichia coli K-12* genome, as computed by Kuiper's statistics. The black horizontal line in panel C shows the minimal clustering above which the essential genes in a genome are considered significantly clustered (i.e. above which the P-value of the Kuiper's test is below 0.05). Where genome "growth" includes gene duplication, the numbers of essential genes is lowered (panel B), and the minimum clustering threshold increases (panel C).

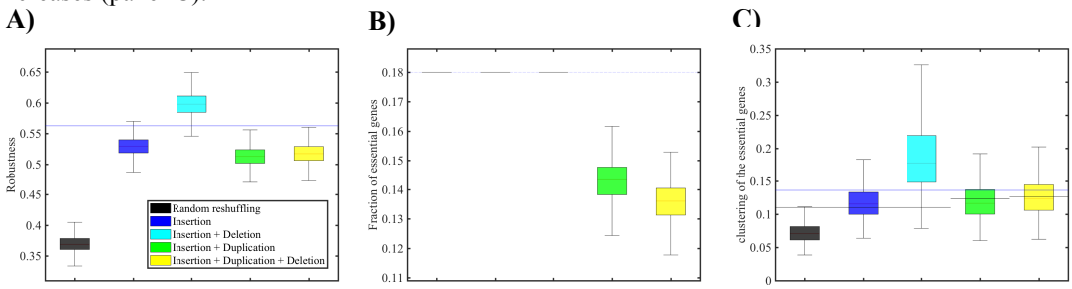

Figure S17: Emergence of essential gene clusters through gradual genomic expansion from a minimal towards a full-sized genome (using randomly reshuffled minimal genomes). Data is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Escherichia coli K-12 MG1655 (iJO1366)*, towards a final genome of equal size as the wild-type *Escherichia coli K-12* genome. Importantly, in this analysis, we have reshuffled the relative ordering of the genes in the minimal genome. We simulated genome "growth" in four different ways (see legend): insertion of genes alone (blue), insertion + deletion (cyan), insertion + duplication (green), and insertion + duplication + deletion (yellow). As a control we generated 100 genomes obtained by random reshuffling of the wild-type *Escherichia coli K-12* genome (black). Vertical axes indicate **A)** robustness to tandem deletions of five genes, **B)** fractions of essential genes, and **C)** clustering of essential genes in the final 100 full-sized genomes, as indicated by the Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line in panel A) indicates metabolic robustness of the wild type *Escherichia coli K-12* genome to tandem deletions of five metabolic genes in glucose minimal medium. The blue horizontal line in C) shows the clustering of essential metabolic genes in the wild type *Escherichia coli K-12* genome, as computed by Kuiper's statistics. The black horizontal line in panel C shows the minimal clustering above which the essential genes in a genome are considered significantly clustered (i.e. above which the P-value of the Kuiper's test is below 0.05). Where genome "growth" includes gene duplication, the number of essential genes is lowered (panel B), and the minimum clustering threshold increases (panel C).

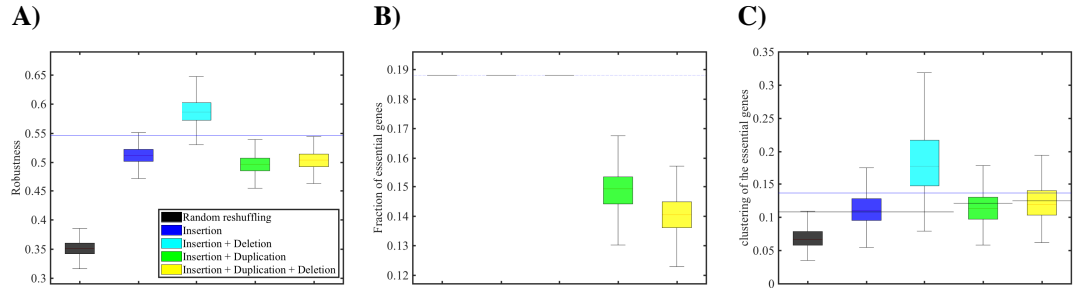

Figure S18: Emergence of essential gene clusters through gradual genomic expansion from a minimal towards a full-sized genome (using acetate as the carbon source). Data is based on 100 genomes "grown" from a minimal genome that is viable on acetate and derived from *Escherichia coli K-12 MG1655 (iJO1366)*, towards a final genome of equal size as the wild-type *Escherichia coli K-12* genome. We simulated genome "growth" in four different ways (see legend): insertion of genes alone (blue), insertion + deletion (cyan), insertion + duplication (green), and insertion + duplication + deletion (yellow). As a control we generated 100 genomes obtained by random reshuffling of the wild-type *Escherichia coli K-12* genome (black). Vertical axes indicate **A)** robustness to tandem deletions of five genes, **B)** fractions of essential genes, and **C)** clustering of essential genes in the final 100 full-sized genomes, as indicated by the Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line in panel A) indicates metabolic robustness of the wild type *Escherichia coli K-12* genome to tandem deletions of five metabolic genes in acetate minimal medium. The blue horizontal line in C) shows the clustering of essential metabolic genes in the wild type *Escherichia coli K-12* genome, as computed by Kuiper's statistics. The black horizontal line in panel C shows the minimal clustering above which the essential genes in a genome are considered significantly clustered (i.e. above which the P-value of the Kuiper's test is below 0.05). Where genome "growth" includes gene duplication, the number of essential genes is lowered (panel B), and the minimum clustering threshold increases (panel C).

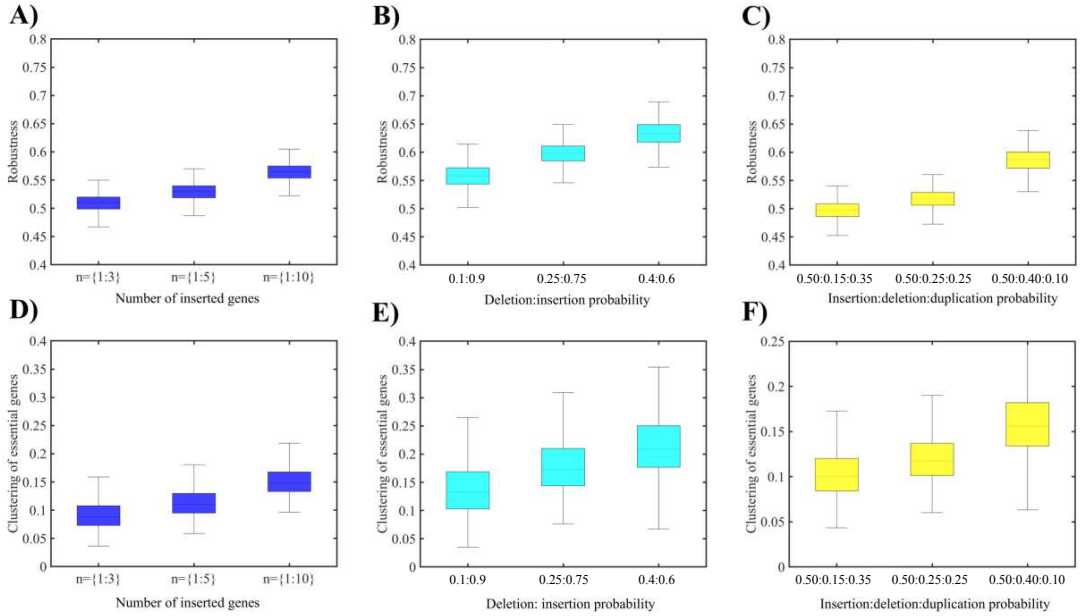

Figure S19: Dependence of the robustness and clustering of essential genes to simulation parameters. Data is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Escherichia coli K-12 MG1655 (iJO1366)*, towards a final genome of equal size as the wild-type *Escherichia coli K-12* genome. In panels **A)** and **D)** we simulated genome "growth" using insertion alone and varied the range of the number of genes (n, horizontal axis) inserted per insertion event. In panels **B)** and **E)** we simulated genome "growth" using insertion and deletion with a varying deletion to insertion ratio (horizontal axis). In panels **C)** and **F)** we simulated genome growth using insertion, deletion and duplication with varying insertion to deletion to duplication ratios. Vertical axes indicate robustness to tandem deletions of five genes (in panels **A-C)**), and clustering of essential genes in the final 100 full-sized genomes, as indicated by Kuiper's test statistic (in panels **D-F)**). Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima.

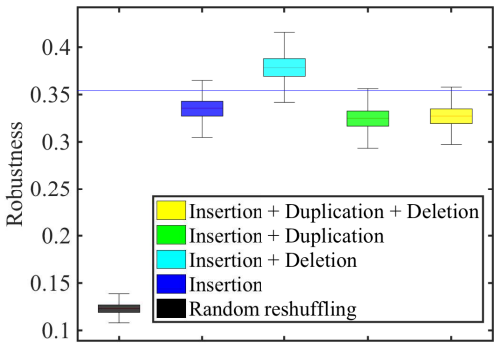

Figure S20: Robustness of the genomes resulting from genome growth simulations to tandem deletions of ten genes. Data is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Escherichia coli K-12 MG1655 (iJO1366)*, towards a final genome of equal size as the wild-type *Escherichia coli K-12* genome. We simulated genome "growth" in four different ways (see legend): insertion of genes alone (blue), insertion + deletion (cyan), insertion + duplication (green), and insertion + duplication + deletion (yellow). As a control, we generated 100 genomes obtained by random reshuffling of the wild-type *Escherichia coli K-12* genome (black). Vertical axes indicate robustness to tandem deletions of ten genes. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima.

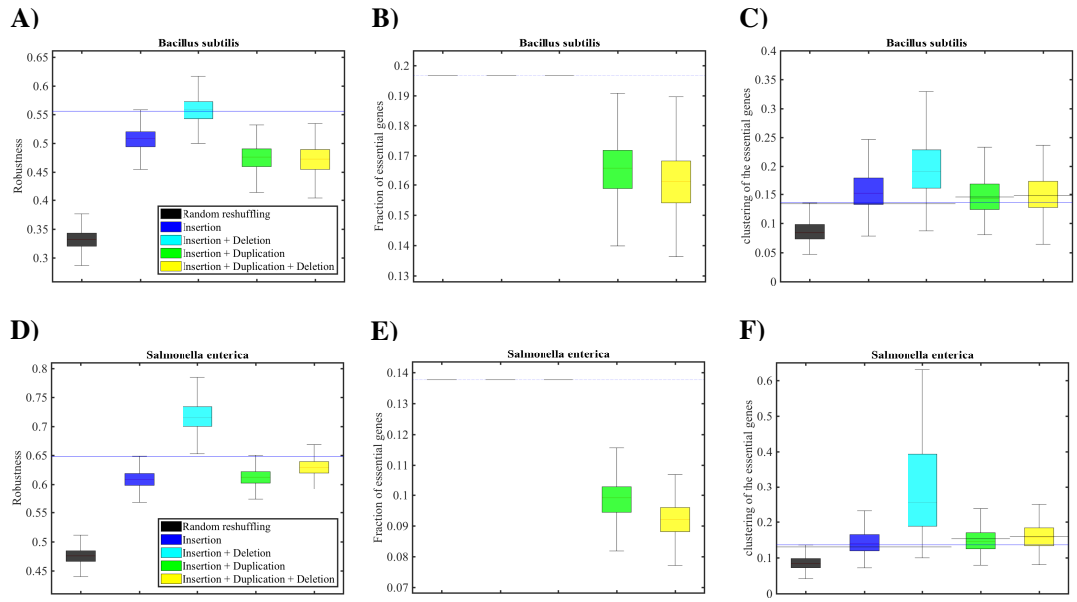

Figure S21: Emergence of essential gene clusters through gradual genomic expansion from a minimal towards a full-sized genome. Data in panels A-C is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Bacillus subtilis*, towards a final genome of equal size as the wild-type *Bacillus subtilis* genome, and the data in panels D-F is based on 100 genomes "grown" from a minimal genome that is viable on glucose and derived from *Salmonella enterica*, towards a final genome of equal size as the wild-type *Salmonella enterica* genome. We simulated genome "growth" in four different ways (see legend): insertion of genes alone (blue), insertion + deletion (cyan), insertion + duplication (green), and insertion + duplication + deletion (yellow). As a control we generated 100 genomes obtained by random reshuffling of the wild-type genome (black). Vertical axes indicate in panels **A)** and **D)** robustness to tandem deletions of five genes, in panels **B)** and **E)** fractions of essential genes, and in panels **C)** and **F)** clustering of essential genes in the final 100 full-sized genomes, as indicated by the Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line in panels A and D indicates metabolic robustness of the wild type *Bacillus subtilis* and *Salmonella enterica* genome to tandem deletions of five metabolic genes in glucose minimal medium. The blue horizontal line in panels C and F shows the clustering of essential metabolic genes in the wild type *Escherichia coli K-12* genome, as computed by Kuiper's statistics. The black horizontal line in panels C and F shows the minimal clustering above which the essential genes in a genome are considered significantly clustered (i.e. above which the P-value of the Kuiper's test is below 0.05). Where genome "growth" includes gene duplication, the number of essential genes is lowered (panels B and E), and the minimum clustering threshold increases (panels C and F).

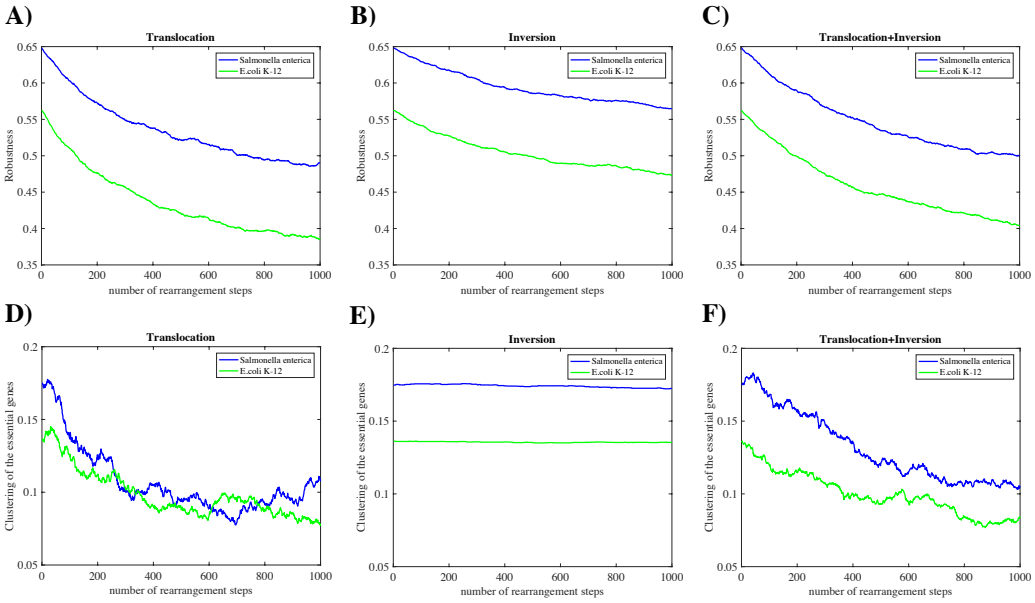

Figure S22: Genome rearrangement can reduce deletional robustness by disrupting the clusters of essential genes (on glucose). In each panel, the horizontal axis shows the number of steps in a simulated genome rearrangement process applied independently to 100 initial genomes derived from the wild-type genomes of two organisms (see legend). In each step, each genome is subjected to a genome rearrangement event (translocation (panels A and D), inversion (panels B and E), and translocation or inversion (panels C and F); see methods). The vertical axes in panels A-C show the average robustness to tandem deletions of five genes. In panels D-F they show the average clustering of essential metabolic genes, as computed by Kuiper's statistics, averaged over all 100 genomes. All simulation data reported are based on minimal media containing glucose as the sole carbon source.

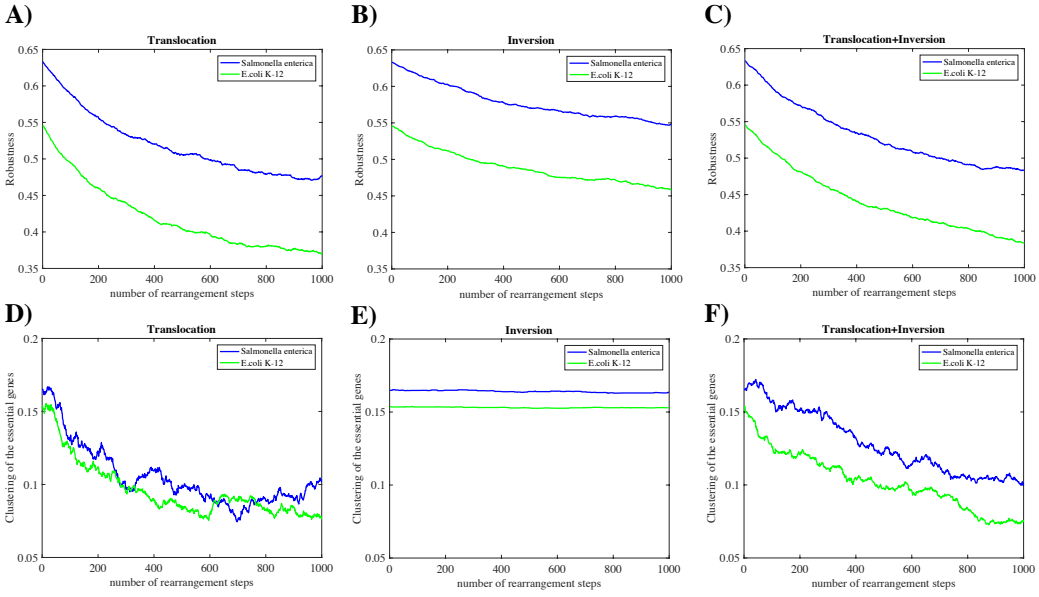

Figure S23: Genome rearrangement can reduce deletional robustness by disrupting the clusters of essential genes (on acetate). In each panel, the horizontal axis shows the number of steps in a simulated genome rearrangement process applied independently to 100 initial genomes derived from the wild-type genomes of two organisms (see legend). In each step, each genome is subjected to a genome rearrangement event (translocation (panels A and D), inversion (panels B and E), and translocation or inversion (panels C and F); see methods). The vertical axes in panels A-C show the average robustness to tandem deletions of five genes, and in panels D-F they show the average clustering of essential metabolic genes, as computed by Kuiper's statistics, averaged over all 100 genomes. All simulation data reported are based on minimal media containing acetate as the sole carbon source.

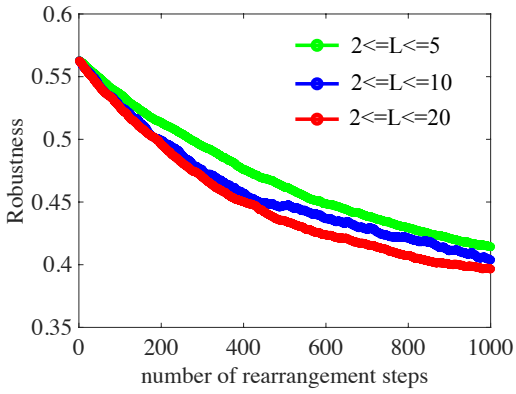

Figure S24: Dependence of robustness reduction through genome rearrangement on the number of rearranged genes. The horizontal axis shows the number of steps in a simulated genome rearrangement process applied independently to 100 initial genomes derived from the wild-type genomes of two organisms (see legend). In each step, each genome is subjected to a genome rearrangement event including translocation or inversion of *n* genes, where *n* varies from i) 2 to 5 (the green curve), ii) 2 to 10 (the blue curve) and iii) 2 to 20 (the red curve). The number of genes chosen for each rearrangement event is chosen with a uniform distribution in each of the indicated intervals. The vertical axis shows the average robustness to tandem deletions of five genes, averaged over all 100 genomes. All simulation data reported are based on minimal media containing glucose as the sole carbon source.

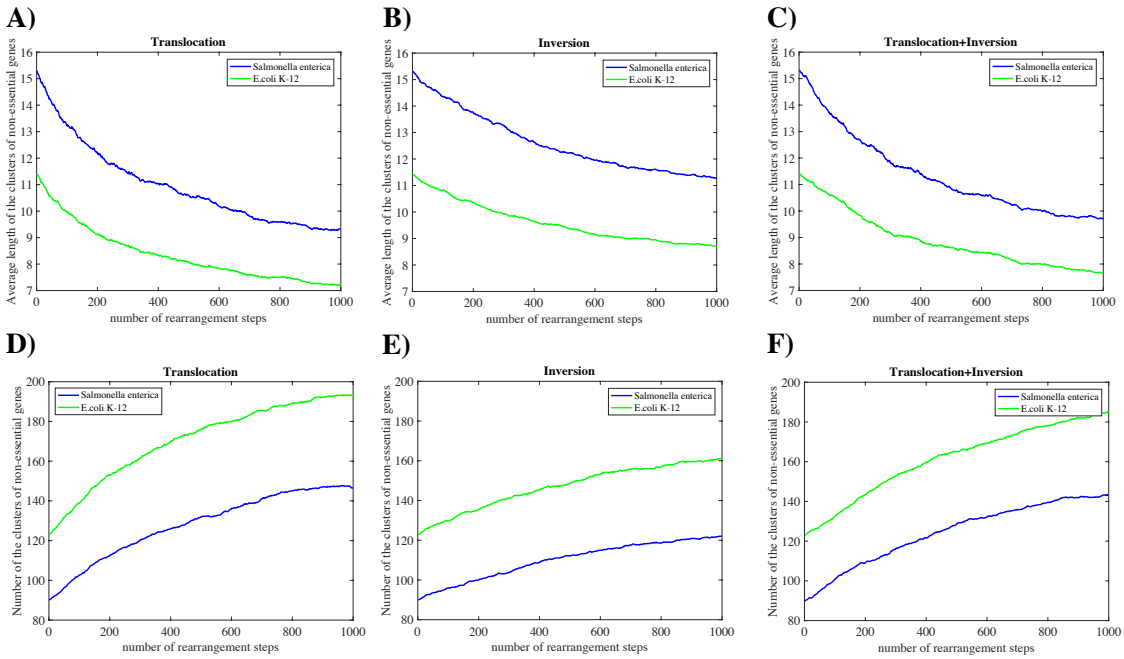

Figure S25: Genome rearrangement can reduce deletional robustness by shrinking the clusters of non-essential genes (on glucose). In each panel, the horizontal axis shows the number of steps in a simulated genome rearrangement process applied independently to 100 initial genomes derived from the wild-type genomes of two organisms (see legend). In each step, each genome is subjected to a genome rearrangement event (translocation (panels **A** and **D**), inversion (panels **B** and **E**), and translocation or inversion (panels **C** and **F**); see methods). The vertical axes in panels A-C show the average length of the clusters of non-essential genes. In panels D-F they show the number of the clusters of non-essential genes, averaged over all 100 genomes. All simulation data reported are based on minimal media containing glucose as the sole carbon source. In this analysis, a cluster of non-essential genes are defined as a set of metabolic genes that are not essential on glucose and intervene between two nearest metabolic genes that are essential on glucose.

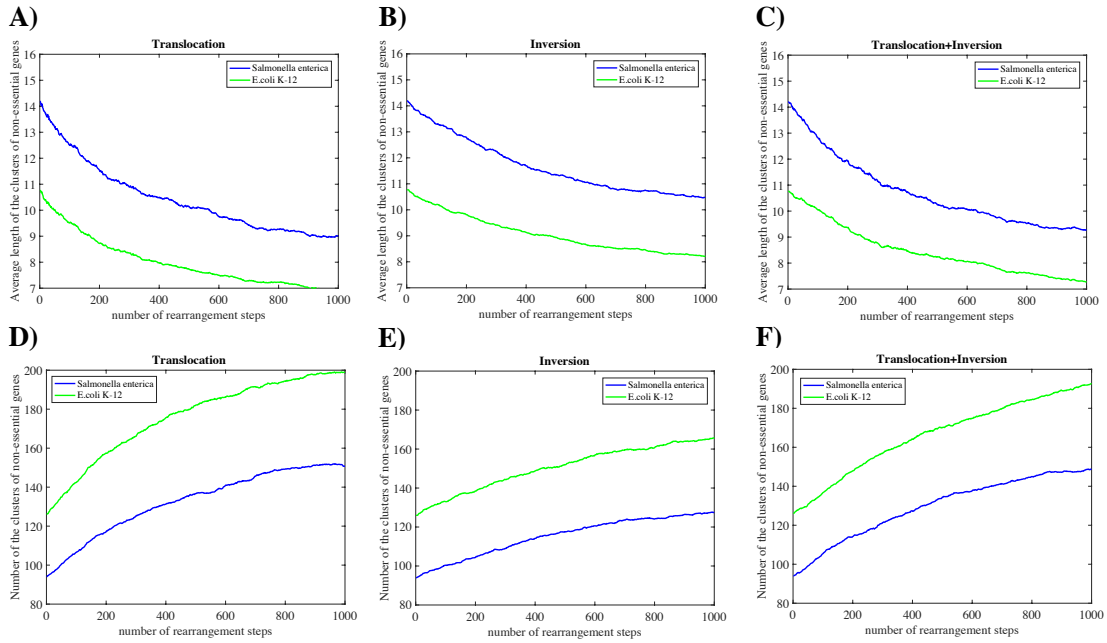

Figure S26: Genome rearrangement can reduce deletional robustness by shrinking the clusters of non-essential genes (on acetate). In each panel, the horizontal axis shows the number of steps in a simulated genome rearrangement process applied independently to 100 initial genomes derived from the wild-type genomes of two organisms (see legend). In each step, each genome is subjected to a genome rearrangement event (translocation (panels **A** and **D**), inversion (panels **B** and **E**), and translocation or inversion (panels **C** and **F**); see methods). The vertical axes in panels A-C show the average length of the clusters of non-essential genes. In panels D-F show the number of the clusters of non-essential genes, averaged over all 100 genomes. All simulation data reported are based on minimal media containing acetate as the sole carbon source. In this analysis, a cluster of non-essential genes are defined as a set of metabolic genes that are not essential on acetate and intervene between two nearest metabolic genes that are essential on acetate.

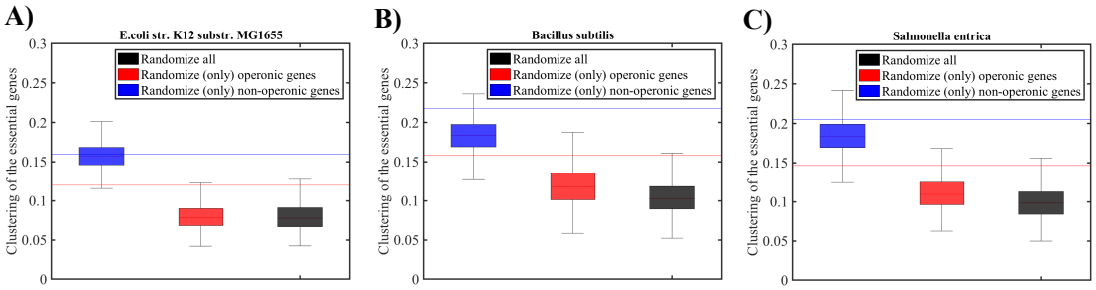

Figure S27: Impact of operonic genes on the clustering of essential genes. Data in this figure are based on partially or completely randomized **A)** from *Escherichia coli K-12 MG1655 (iJO1366),* **B)** *Salmonella enterica* genome, and **C)** *Bacillus subtilis* genome, with randomized orders of i) all genes (black), ii) non-operonic genes (blue), iii) operonic genes (red). The vertical axes show the extent of clustering of strictly essential genes in the 100 randomized genomes, as indicated by the Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line indicates the clustering of strictly essential genes in the corresponding wild type genomes as indicated by the Kuiper's test statistic, and the red horizontal line shows a minimal threshold of significant clustering (i.e. where Kuiper's test yields $P<0.05$).

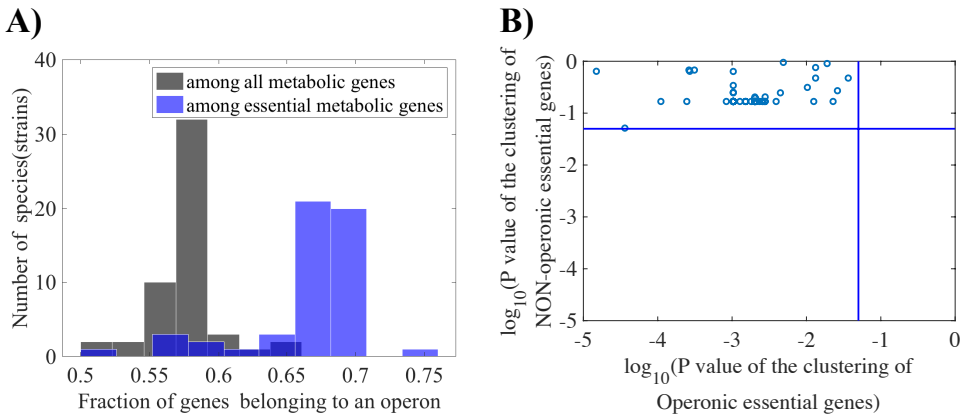

Figure S28: Operons and essential genes. A) Histogram of the fraction of all metabolic genes (black), and the fraction of strictly essential metabolic genes (blue) which belong to an operon, based on the 52 species or strains used in this analysis. We consider a metabolic gene as strictly essential, if its deletion results in losing viability on all the carbon sources on which the wild type metabolism is viable. **B)** In this analysis, we subdivided all strictly essential genes in each of 52 metabolic genome into two groups i) those belonging to an operon and ii) those not belonging to an operon. For each genomes, we determined the extent of gene clustering using the *P*-value generated by Kuiper's test. P-values are adjusted for multiple-testing using the Benjamini-Hochberg's correction (32). Each circle in this figure corresponds to a given species or strain. The horizontal and vertical axes show the extent of clustering for genes that are part of an operon and not part of an operon, respectively. The blue lines correspond to a significance threshold of P=0.05 ($-log_{10}0.05$). Note that operonic essential genes are significantly clustered in all genomes. Whereas none of our 52 genomes show evidence for clustering of essential genes outside operons (at P=0.05), essential genes in operons are clustered in all 52 genomes.

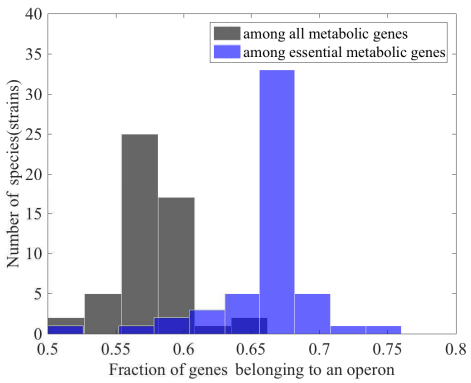

Figure S29: Operons and conditionally essential genes. Histogram of the fraction of all metabolic genes (black), and the fraction of conditionally essential metabolic genes (blue) which belong to an operon among the 52 species (strains) used in this analysis. We consider a metabolic gene as conditionally essential, if its deletion results in losing viability on at least one of the carbon sources on which the wild type metabolism is viable.

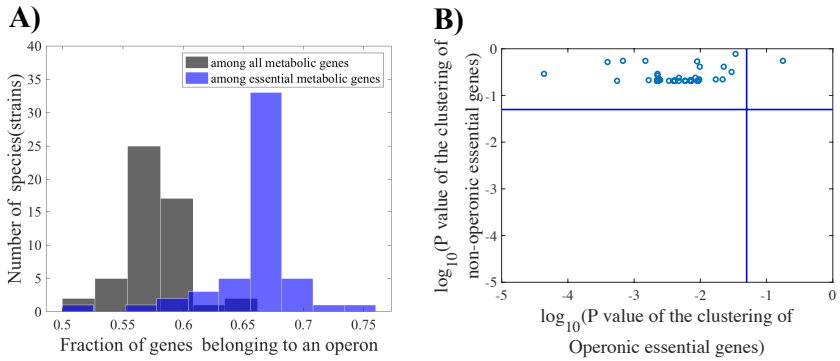

Figure S30: Operons and the essential genes (on glucose as minimal media). A) Histogram of the fraction of metabolic genes (black), and the fraction of essential metabolic genes (blue) which belong to an operon among the 52 species (strains) used in this analysis. We consider a metabolic gene as essential, if its deletion results in losing viability on glucose. **B)** In this analysis, we subdivided all strictly essential genes in each of 52 metabolic genomes into two groups i) those belonging to an operon and ii) those not belonging to an operon. For each genome, we determined the extent of gene clustering using the *P*-value generated by Kuiper's test. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). Each circle in this figure corresponds to a given species or strain. The horizontal and vertical axes show the extent of clustering for essential genes that are part of an operon and not part of an operon, respectively. The blue lines correspond to a significance threshold of P=0.05 ($-log_{10}0.05$). Note that operonic essential genes are significantly clustered in more genomes. Whereas none of the 52 genomes show evidence for clustering of essential genes outside operons (at P=0.05), 51 of the 52 genomes (98.08%) show clustering of essential genes in operons.

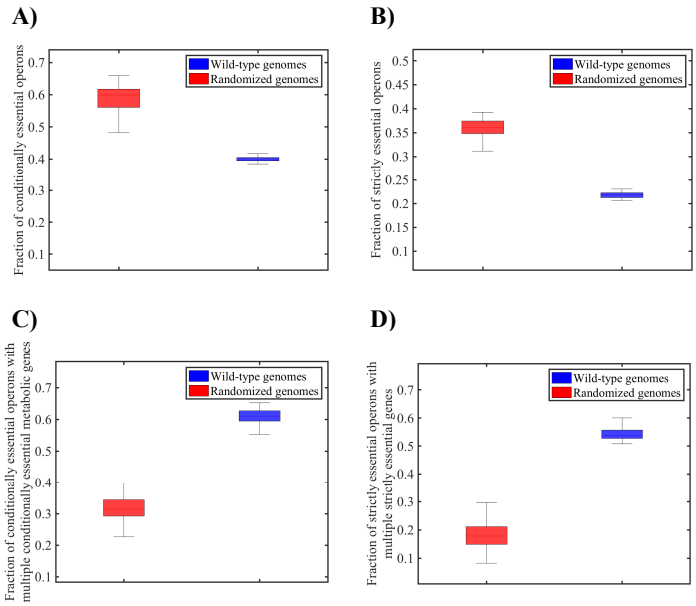

Figure S31: Overabundance of operons with multiple essential genes. Horizontal axes show the fraction of **A)** conditionally essential operons (i.e. operons with at least one conditionally essential metabolic gene), **B)** strictly essential operons (i.e. operons with multiple conditionally essential metabolic genes), **C)** conditionally essential operons with multiple conditionally essential metabolic genes, and **D)** strictly essential operons with multiple strictly essential metabolic genes. All data are based on 52 wild type (blue) and randomized (red) bacterial genomes. In this analysis, only operonic genes are randomized, such that the total number of essential metabolic genes is the same in wild-type genomes and in randomized genomes. However, operonic essential genes are uniformly distributed among operons in randomized genomes. Thus, a higher fraction of operons are essential in such genomes, but a lower fraction of essential operons contain multiple essential genes than in wild-type genomes. Moreover, the lower fraction of essential operons (caused by the overabundance of essential operons with multiple essential genes) in wild-type genomes explains the higher robustness to operon deletions compared to tandem deletions of the same size that we observed in our analysis pertaining to the selfish operon hypothesis (See text S11 and figure S45).

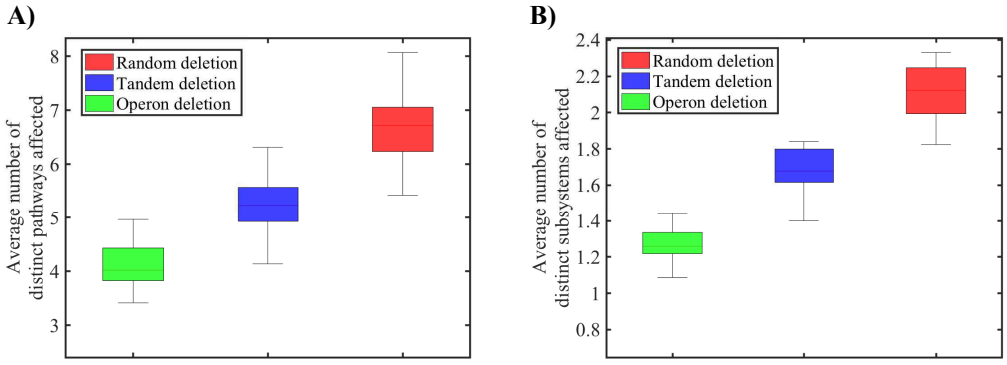

Figure S32: Deletion of an operon affects fewer metabolic pathways and functional subsystems than tandem deletion of the same number of non-operonic genes. Boxplots of the average number of **A)** metabolic pathways and **B)** functional subsystems affected by operon deletion ($\overline{NO}_P$ and $\overline{NO}_{sub}$ green), as well as by tandem ($\overline{NT}_P$ and $\overline{NT}_{sub}$ blue), and random deletion ($\overline{NR}_P$ and $\overline{NR}_{sub}$ red), of the same number of metabolic genes, computed for 52 bacterial genomes. $\overline{NO}_P$ is significantly smaller than $\overline{NT}_P$ (Paired-sample t-test: P-value<10^{-16}) and $\overline{NR}_P$ (Paired-sample t-test: P-value<10^{-31}). Similarly, $\overline{NO}_{sub}$ is significantly smaller than $\overline{NT}_{sub}$ (Paired-sample t-test: P-value<10^{-47}) and $\overline{NR}_{sub}$ (Paired-sample t-test: P-value<10^{-42}). Thus, metabolic genes belonging to the same operon tend to be a part of the same metabolic pathway or functional subsystem. This also causes $\overline{NT}_P$ to be significantly smaller than $\overline{NR}_P$ (Paired-sample t-test: P-value<10^{-19}). Similarly, $\overline{NT}_{sub}$ is significantly smaller than $\overline{NR}_{sub}$ (Paired-sample t-test: P-value<10^{-36}). See text S8 for more details.

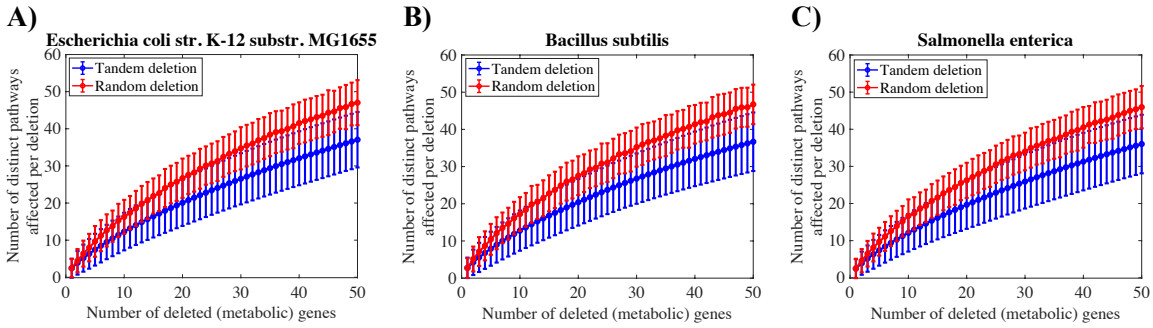

Figure S33: Random deletion affects more metabolic pathways than tandem deletion. Data in each panel show the mean (circles) and standard deviation (vertical bars) of the number of distinct metabolic pathways affected by tandem (blue) and random (red) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

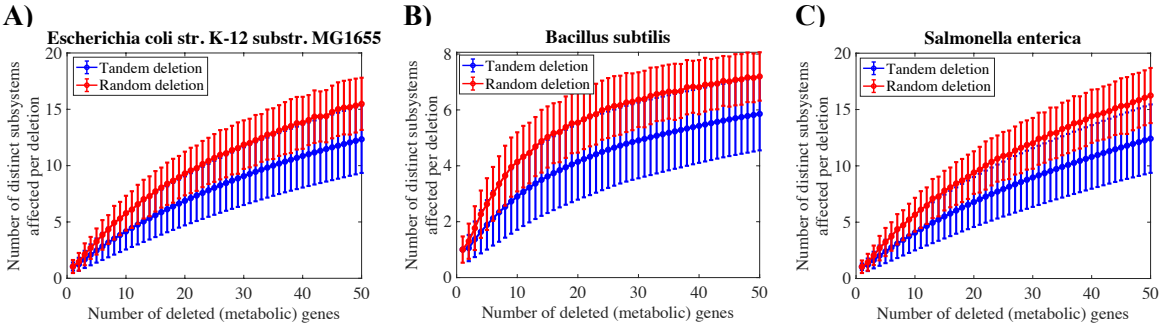

Figure S34: Random deletion affects more functional subsystems than tandem deletion. Data in each panel show the mean (circles) and standard deviation (vertical bars) of the number of distinct functional subsystems affected by tandem (blue) and random (red) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

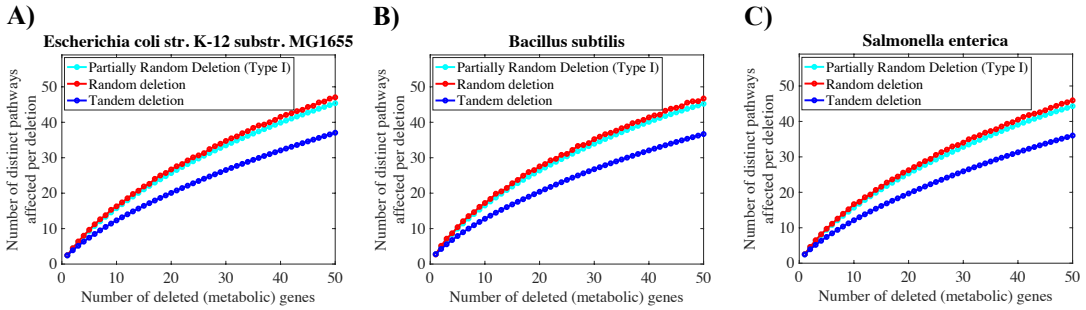

Figure S35: The number of metabolic pathways affected by partially random deletions (type1) is considerably higher than that affected by tandem deletion. Data in each panel show the mean of the number of distinct metabolic pathways affected by tandem (blue), random (red) and partially random (type I; cyan) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

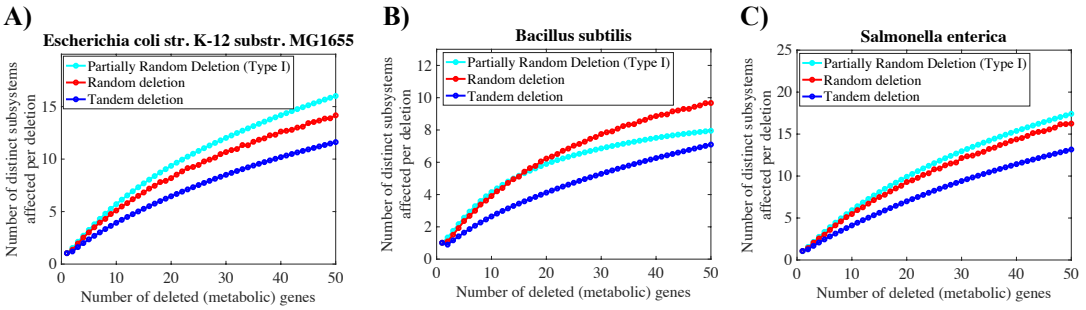

Figure S36: The number of functional subsystems affected by partially random deletions (type1) is considerably higher than that affected by tandem deletion. Data in each panel show the mean of the number of distinct functional subsystems affected by tandem (blue), random (red) and partially random (type I; cyan) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

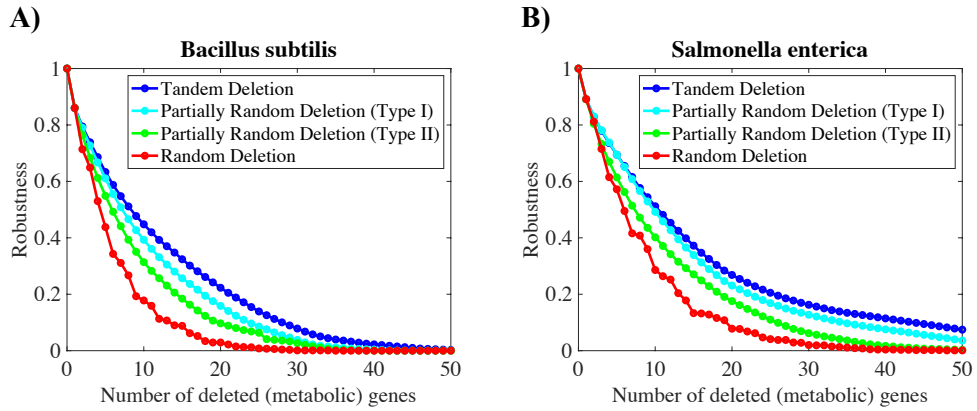

Figure S37: De-convolution of the effect of the clustering of essential genes versus the number of affected metabolic pathways on the excess robustness to tandem deletion. Data in each panel show the robustness to tandem (blue), random (red), partially random (cyan and green for type I and II respectively) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Bacillus subtillis* and **B)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

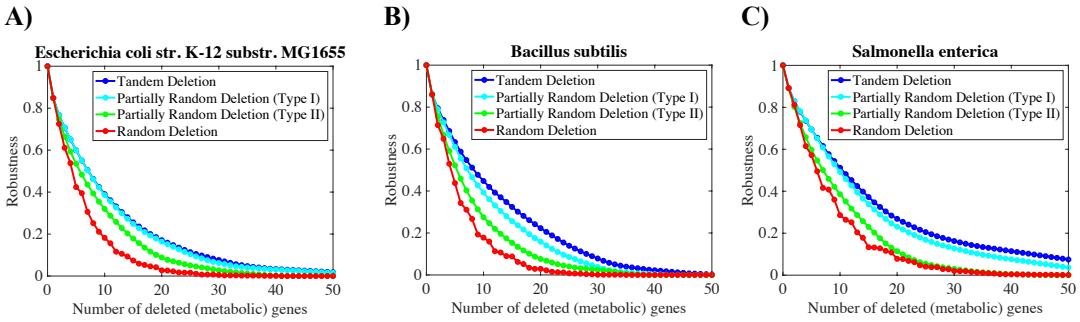

Figure S38: De-convolution of the effect of the clustering of essential genes versus the number of affected functional subsystems on the excess robustness to tandem deletion. Data in each panel show the robustness to tandem (blue), random (red), partially random (cyan and green for type I and II respectively) deletion of a given number of metabolic genes (horizontal axis) in **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

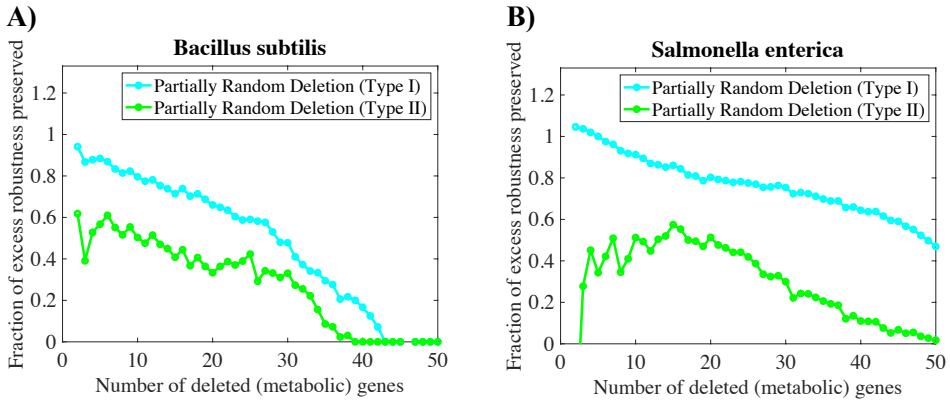

Figure S39: The effect of the organization of essential genes is consistently higher than the effect of the organization of metabolic pathways on the excess robustness to tandem deletion. Data in each panel show the fraction of excess robustness to tandem deletion that is preserved after partially random deletion of type I (cyan) and type II (green), as a function of the number of deleted metabolic genes (horizontal axis) for **A)** *Bacillus subtillis* and **B)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

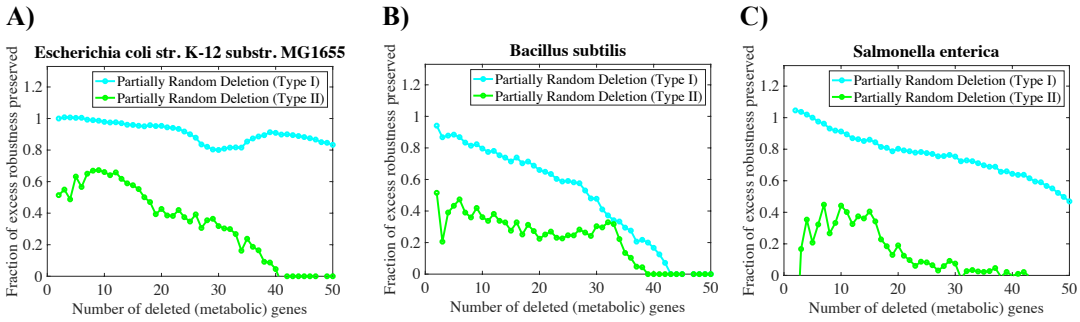

Figure S40: The effect of the organization of essential genes is consistently higher than the effect of the organization of functional subsystems on the excess robustness to tandem deletion. Data in each panel show the fraction of excess robustness to tandem deletion that is preserved after partially random deletion of type I (cyan) and type II (green), as a function of the number of deleted metabolic genes (horizontal axis) for **A)** *Escherichia coli K-12 MG1655 (iJO1366)*, **B)** *Bacillus subtillis* and **C)** *Salmonella enterica* genome. Each curve is obtained by a linear interpolation between the 50 data points. See text S8 for more details.

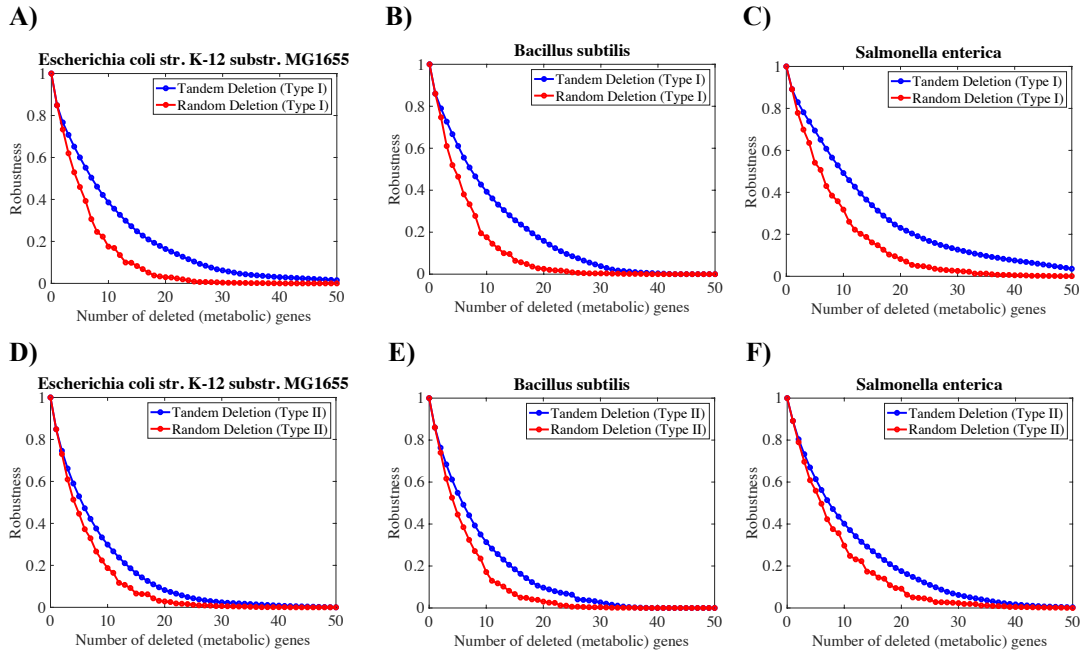

Figure S41: Robustness to tandem deletion versus random deletion in partially randomized genomes (Types I and II). In each panel, the vertical axis shows the robustness to tandem (blue) and random (red) deletion of metabolic genes, averaged over all deletional variants we examined, as a function of the number of deleted genes (horizontal axis). Upper panels **(A-C)** correspond to partially randomized genomes of type I, whereas the lower panels **(D-E)** correspond to the randomized genomes of type II. The partially randomized genomes are obtained from their corresponding wild-type genomes (**A, D)** *Escherichia coli K-12 MG1655 (iJO1366)*, (**B, E)** *Bacillus subtillis* and (**C, F)** *Salmonella enterica*. While the robustness to tandem deletion in partially randomized genomes of type I (upper panels) is substantially higher than the robustness to random deletion, in partially randomized genomes of type II (lower panels) the difference between tandem and random robustness is not very large. Note that in partially randomized genomes of type I the organization of essential genes remains the same as that of wild-type genome, while in partially randomized genomes of type II the organization of metabolic pathways are kept the same as in the wild-type genome. Interpolation between data points is linear and is displayed as a visual guide.

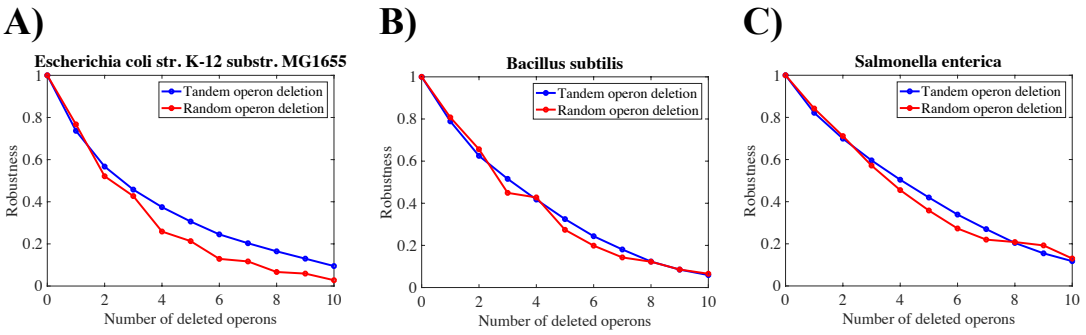

Figure S42: Robustness to tandem operon deletion versus random operon deletion (conditional definition of robustness). The vertical axes show the robustness of *Escherichia coli K-12 MG1655 (iJO1366)* (panel **A**), *Salmonella enterica* (panel **B**) and *Bacillus subtillis* (panel **C**) to tandem (blue) and random (red) deletion of operons, averaged over all deletional variants we examined, as a function of the number of deleted operons (horizontal axis). Robustness is defined conditionally, i.e., as the fraction of deletional variants that retain viability on at least one carbon source. Interpolation between data points is linear and is displayed as a visual guide.

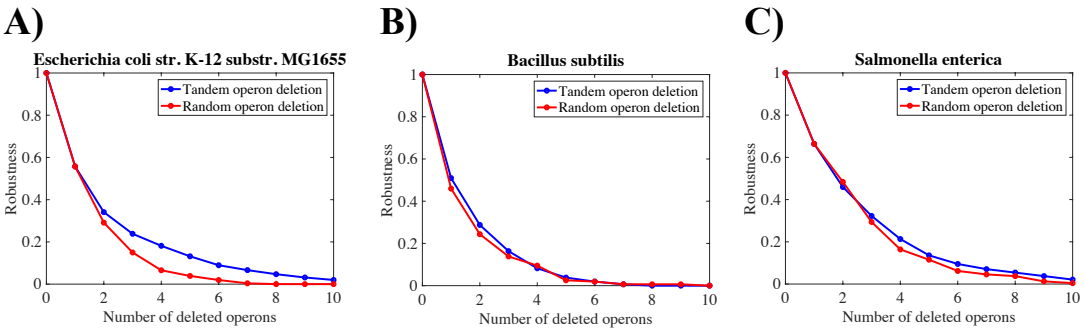

Figure S43: Robustness to tandem operon deletion versus random operon deletion (strict definition of robustness). The vertical axes show the robustness of *Escherichia coli K-12 MG1655 (iJO1366)* (panel **A**), *Salmonella enterica* (panel **B**) and *Bacillus subtillis* (panel **C**) to tandem (blue) and random (red) deletion of operons, averaged over all deletional variants we examined, as a function of the number of deleted operons (horizontal axis). Robustness is defined strictly, i.e., as the fraction of deletional variants that retain viability on all carbon sources on which the wild type metabolism is viable. Interpolation between data points is linear and is displayed as a visual guide.

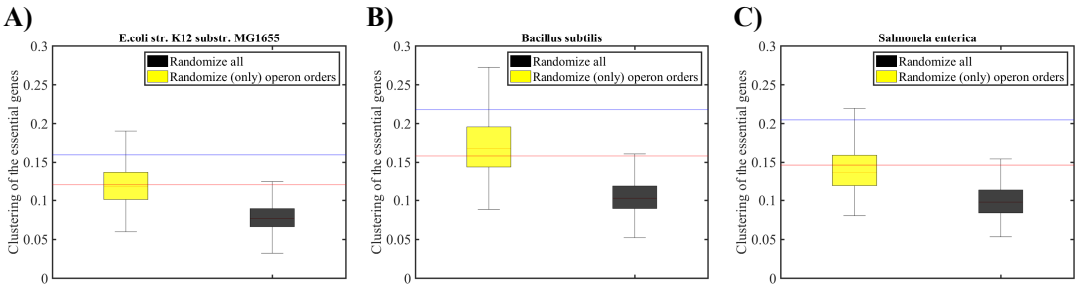

Figure S44: Impact of operon orders on the clustering of essential genes. Data in this figure are based on partially or completely randomized **A)** from *Escherichia coli K-12 MG1655 (iJO1366),* **B)** *Salmonella enterica* genome, and **C)** *Bacillus subtilis* genome, with randomized orders of i) all genes (black), ii) entire operons (yellow, without changing the intra-operonic gene orders). The vertical axes show the extent of clustering of strictly essential genes in the 100 randomized genomes, as indicated by the Kuiper's test statistic. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. The blue horizontal line indicates the clustering of strictly essential genes in the corresponding wild type genomes as indicated by the Kuiper's test statistic, and the red horizontal line shows a minimal threshold of significant clustering (i.e. where Kuiper's test yields P<0.05).

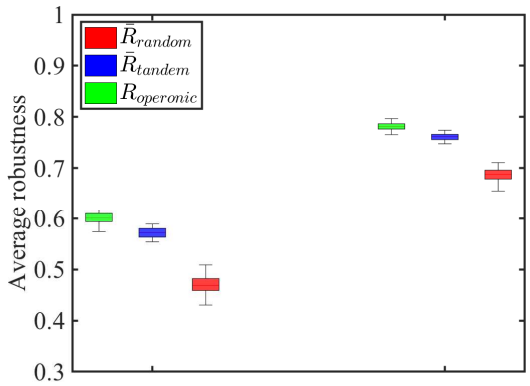

Figure S45: Robustness to operon deletion. Box plots of robustness to operon deletion ($R_{Operonic}$, blue) and of average robustness to tandem ($\bar{R}_{tandem}$, blue) and random ($\bar{R}_{random}$, red) deletion of the same length (see text S11), computed for 52 bacterial species. For the box plots of the left-hand side, robustness is defined based on retaining viability on all carbon sources on which the wild-type genome is viable, while for the boxplots of the right hand side, robustness is defined based on retaining viability on at least one carbon sourcE.

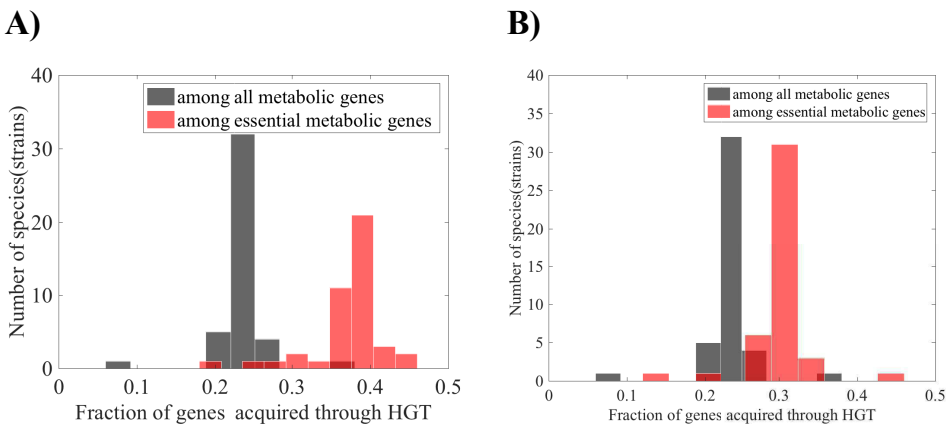

Figure S46: HGT and the clustering of essential genes. A) Histogram of the fraction of metabolic genes (black), and the fraction of strictly essential metabolic genes (red) which have been acquired through horizontal gene transfer (HGT) among the 43 species (strains) used in this analysis. We consider a metabolic gene as strictly essential, if its deletion results in losing viability on all the carbon sources on which the wild type metabolism is viable. **B)** Histogram of the fraction of metabolic genes (black), and the fraction of conditionally essential metabolic genes (red) which have been acquired through horizontal gene transfer (HGT) among the 43 species (strains) used in this analysis. We consider a metabolic gene as conditionally essential, if its deletion results in losing viability on at least one of the carbon sources on which the wild type metabolism is viable.

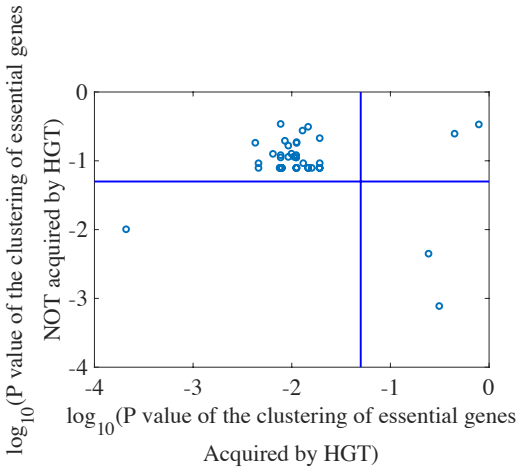

Figure S47: HGT and the clustering of essential genes. In this analysis, we subdivided all strictly essential genes in each metabolic genome into two groups: i) those acquired by horizontal gene transfer (HGT) and ii) those not acquired by horizontal gene transfer. For each of the 43 genomes, and separately for strictly essential genes in each of the two groups, we determined the extent of gene clustering using the P-value generated by Kuiper's test. P-values are adjusted for multiple-testing using the Benjamini-Hochberg approach (32). Each circle in this figure corresponds to a given species or strain. The horizontal and vertical axes show the extent of clustering for genes acquired and not acquired by HGT, respectively. The blue lines correspond to a significance threshold of P=0.05 ($-log_{10}0.05$). Note that horizontally transferred genes show greater clustering in the vast majority of genomes. Whereas the clustering of horizontally transferred genes is significant at P=0.05 in 39 among 43 genomes (90.7%), that of not horizontally transferred genes is significant only in 3 genomes (6.98%).

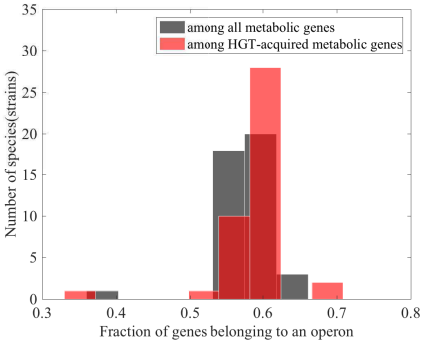

Figure S48: Horizontal gene transfer and operons. A) Histogram of the fraction of metabolic genes (black), and the fraction of HGT-acquired metabolic genes (red), which belong to an operon among the 43 species (strains) used in this analysis. The two distributions overlap to a great extent, indicating that HGT-acquired genes do not have substantially greater propensity to be located in operons than non HGT-acquired genes.

Supplementary tables:

Index	Carbon source	Index	Carbon source
1	D-Glucose	52	L-Tryptophan
2	Uracil	53	Maltose
3	Acetoacetate	54	L-Asparagine
4	3-(3-hydroxy-phenyl)propionate	55	L-Lactate
5	N-Acetyl-D-glucosamine	56	(S)-Propane-1
6	Acetaldehyde	57	D-Ribose
7	N-Acetyl-D-mannosamine	58	Sucrose
8	L-Cysteine	59	Thymidine
9	2-Dehydro-3-deoxy-D-gluconate	60	D-serine
10	Tetradecanoate (n-C14:0)	61	D-Galactose
11	N-Acetylneuraminate	62	Lactose
12	L-Glutamate	63	L-Malate
13	Uridine	64	L-Aspartate
14	Xanthine	65	Putrescine
15	L-Arginine	66	D-Glucose 6-phosphate
16	L-Alanine	67	Phenylpropanoate
17	Glycolate	68	Butyrate (n-C4:0)
18	Guanine	69	Octadecanoate (n-C18:0)
19	Glycine	70	Trehalose
20	4-Aminobutanoate	71	L-Histidine
21	L-Glutamine	72	Pyruvate
22	Adenine	73	D-Mannitol
23	Guanosine	74	Citrate
24	Glycerol 3-phosphate	75	L-tartrate
25	D-Glucuronate	76	L-Threonine
26	Glycerol	77	Ornithine
27	Hexadecanoate (n-C16:0)	78	Maltopentaose
28	Adenosine	79	Maltotriose
29	D-Glyceraldehyde	80	Maltohexaose
30	D-Glucosamine	81	L-Rhamnose
31	D-Galacturonate	82	Succinate
32	D-Galactonate	83	D-Mannose
33	D-Glucarate	84	Cytidine
34	Hypoxanthine	85	D-Sorbitol
35	D-Gluconate	86	Deoxyadenosine
36	2-Oxoglutarate	87	Maltotetraose
37	Galactitol	88	Melibiose
38	3-hydroxycinnamic acid	89	D-Mannose 6-phosphate
39	Allantoin	90	Deoxycytidine
40	D-Galactarate	91	L-Fucose
41	D-Xylose	92	L-Serine
42	L-Idonate	93	D-Fructose
43	Acetate	94	Deoxyguanosine
44	Xanthosine	95	Dihydroxyacetone
45	AMP	96	Fumarate
46	L-Isoleucine	97	Cytosine
47	Inosine	98	Deoxyinosine
48	L-Arabinose	99	D-Alanine
49	L-Valine	100	Deoxyuridine
50	D-Lactate	101	Formate
51	L-Proline	102	Ethanol

Table S2: List of carbon sources used in this study.

Species (strain)	Number of strictly essential genes	Number of conditionally essential genes
E.coli K-12 MG1655 [iAF1260]	144	322
Methanosarcina barkeri str. Fusaro	139	139
Geobacter metallireducens GS-15	268	268
E. coli APEC O1	160	320
E. coli BL21(DE3) [iB21 1397]	164	326
E. coli BW2952	210	367
E. coli CFT073	204	204
E. coli O127:H6	160	313
E. coli 042	159	329
E. coli 55989	159	326
E. coli ABU 83972	160	320
E. coli B str. REL606	159	327
E. coli BL21-Gold(DE3)pLysS AG	164	326
E. coli BL21(DE3) [iECD1391]	164	327
E. coli DH1 [iEcDH1 1363]	159	323
E. coli DH1 [iECDH1ME8569 1439]	158	322
E. coli E24377A	157	324
E. coli ED1a	161	319
E. coli O157:H7	159	357
E. coli HS	159	336
E. coli NA114	158	320
E. coli O103:H2 str. 12009	159	322
E. coli O111:H- str. 11128	159	346
E. coli O26:H11 str. 11368	159	337
E. coli IHE3034	160	320
E. coli ATCC 8739	159	323
E. coli 536	160	320
E. coli O157:H7 str. Sakai	159	323
E. coli S88	160	319
E. coli SE11	159	323
E. coli SE15	210	375
E. coli SMS-3-5	159	328
E. coli O157:H7 str. TW14359	159	322
E. coli UMN026	158	332
E. coli W [iECW_1372]	158	326
E. coli KO11FL	158	321
E. coli ETEC H10407	159	333
E. coli O55:H7 str. CB9615	159	337
E. coli K-12 MG1655 [iJO1366]	207	383
E. coli K-12 MG1655 [iJR904]	134	316
E. coli LF82	182	403
Mycobacterium tuberculosis H37Rv	197	197
E. coli O83:H1 str. NRG 857C	160	317
Shigella flexneri 2a str. 2457T	167	334
Staphylococcus aureus N315	61	205
Shigella flexneri 5 str. 8401	164	320
E. coli UM146	159	319
E. coli UMNK88	159	319
E. coli UTI89	160	320
E. coli W [iWFL 1372]	158	326
E. coli str. K-12 W3110	159	329
Klebsiella pneumoniae MGH78578	89	402
Bacillus subtilis str. 168	120	288
E. coli O157:H7 str. EDL933	159	322
Salmonella Typhimurium str. LT2	140	293

Table S3: Number of conditionally and strictly essential metabolic genes in bacterial genomes. Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, number of strictly essential metabolic genes, and number of conditionally essential metabolic genes. We consider a metabolic gene strictly essential, if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable, and we consider a metabolic gene conditionally essential, if its deletion abolishes viability on at least one carbon source.

Species (strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
E.coli K-12 MG1655 [iAF1260]	1	6.697778e-03	0.1713	0.1435
Methanosarcina barkeri str. Fusaro	0	2.576852e-01	0.1211	0.1492
Geobacter metallireducens GS-15	1	5.060000e-04	0.1688	0.1078
E. coli APEC O1	1	7.626667e-04	0.2082	0.1371
E. coli BL21(DE3) [iB21 1397]	1	4.809302e-03	0.1656	0.1354
E. coli BW2952	1	1.906667e-03	0.1694	0.1197
E. coli CFT073	1	2.039783e-03	0.1687	0.1242
E. coli O127:H6	1	3.111842e-03	0.1733	0.1371
E. coli 042	1	2.039783e-03	0.1839	0.1375
E. coli 55989	1	2.039783e-03	0.1858	0.1375
E. coli ABU 83972	1	2.218966e-03	0.1792	0.1371
E. coli B str. REL606	1	4.809302e-03	0.1682	0.1375
E. coli BL21-Gold(DE3)pLysS AG	1	7.688043e-03	0.1599	0.1354
E. coli BL21(DE3) [iECD1391]	1	3.616250e-03	0.1692	0.1354
E. coli DH1 [iEcDH1 1363]	1	2.039783e-03	0.1856	0.1375
E. coli DH1 [iECDH1ME8569 1439]	1	2.716327e-02	0.1474	0.138
E. coli E24377A	1	2.039783e-03	0.1855	0.1375
E. coli ED1a	1	6.337500e-03	0.164	0.1367
E. coli O157:H7	1	2.115385e-03	0.1816	0.1375
E. coli HS	1	2.658333e-03	0.1759	0.1375
E. coli NA114	1	3.809756e-03	0.1716	0.138
E. coli O103:H2 str. 12009	1	2.283333e-03	0.1782	0.1375
E. coli O111:H- str. 11128	1	2.039783e-03	0.1828	0.1375
E. coli O26:H11 str. 11368	1	2.039783e-03	0.1833	0.1375
E. coli IHE3034	1	3.257692e-03	0.1726	0.1371
E. coli ATCC 8739	1	2.218966e-03	0.18	0.1375
E. coli 536	1	2.251563e-03	0.1782	0.1371
E. coli O157:H7 str. Sakai	1	2.115385e-03	0.1812	0.1375
E. coli S88	1	3.077027e-03	0.1737	0.1371
E. coli SE11	1	2.039783e-03	0.1852	0.1375
E. coli SE15	1	2.442647e-03	0.1543	0.1197
E. coli SMS-3-5	1	2.039783e-03	0.1835	0.1375
E. coli O157:H7 str. TW14359	1	2.218966e-03	0.1799	0.1375
E. coli UMN026	1	2.251563e-03	0.1785	0.1371
E. coli W [iECW_1372]	1	1.906667e-03	0.196	0.138
E. coli KO11FL	1	2.039783e-03	0.1876	0.138
E. coli ETEC H10407	1	2.039783e-03	0.1843	0.1375
E. coli O55:H7 str. CB9615	1	1.906667e-03	0.192	0.1375
E. coli K-12 MG1655 [iJO1366]	1	2.039783e-03	0.1596	0.12
E. coli K-12 MG1655 [iJR904]	1	3.465000e-02	0.1565	0.1498
E. coli LF82	1	1.133000e-07	0.275	0.1371
Mycobacterium tuberculosis H37Rv	0	2.950000e-01	0.097	0.123
E. coli O83:H1 str. NRG 857C	1	2.498571e-03	0.1763	0.1371
Shigella flexneri 2a str. 2457T	1	1.294792e-02	0.1533	0.135
Staphylococcus aureus N315	0	6.875000e-02	0.2198	0.2255
Shigella flexneri 5 str. 8401	1	1.143298e-02	0.1554	0.1354
E. coli UM146	1	2.039783e-03	0.184	0.1375
E. coli UMNK88	0	1.795283e-01	0.1185	0.1375
E. coli UTI89	1	2.251563e-03	0.1782	0.1371
E. coli W [iWFL 1372]	1	1.906667e-03	0.196	0.138
E. coli str. K-12 W3110	1	2.115385e-03	0.1815	0.1375
Klebsiella pneumoniae MGH78578	1	4.572549e-02	0.1922	0.1887
Bacillus subtilis str. 168	1	2.039783e-03	0.218	0.1583
E. coli O157:H7 str. EDL933	1	1.906667e-03	0.1923	0.1375
Salmonella Typhimurium str. LT2	1	1.906667e-03	0.2048	0.1465

Table S4: Clustering of the strictly essential genes. Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, whether the null hypothesis of a uniform distribution of strictly essential genes is rejected by Kuiper's test (1) or not (0), the P-value of Kuiper's test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 51 among the 55 genomes the null hypothesis is rejected, i.e., essential genes are significantly clustered. We consider a metabolic gene strictly essential if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species (strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
E.coli K-12 MG1655 [iAF1260]	1	7.547222e-06	0.1615	0.0965
Methanosarcina barkeri str. Fusaro	0	3.025000e-01	0.1151	0.146
Geobacter metallireducens GS-15	1	3.126316e-05	0.1644	0.1056
E. coli APEC O1	1	6.927381e-05	0.1456	0.0968
E. coli BL21(DE3) [iB21 1397]	1	2.300000e-05	0.1516	0.0959
E. coli BW2952	1	1.827419e-05	0.1444	0.0904
E. coli CFT073	1	1.833333e-04	0.1743	0.1209
E. coli O127:H6	1	2.068478e-04	0.1402	0.0978
E. coli 042	1	9.350000e-06	0.1579	0.0954
E. coli 55989	1	9.361000e-07	0.1764	0.0959
E. coli ABU 83972	1	1.620667e-05	0.1554	0.0968
E. coli B str. REL606	1	1.283333e-05	0.1555	0.0957
E. coli BL21-Gold(DE3)pLysS AG	1	4.735366e-05	0.1465	0.0959
E. coli BL21(DE3) [iECD1391]	1	1.283333e-05	0.1558	0.0957
E. coli DH1 [iEcDH1 1363]	1	6.256250e-06	0.1629	0.0963
E. coli DH1 [iECDH1ME8569 1439]	1	2.285937e-05	0.1527	0.0965
E. coli E24377A	1	1.987857e-06	0.172	0.0962
E. coli ED1a	1	3.124468e-04	0.1362	0.0969
E. coli O157:H7	1	1.117500e-04	0.1351	0.0917
E. coli HS	1	6.128571e-06	0.1603	0.0945
E. coli NA114	1	4.735366e-05	0.1481	0.0968
E. coli O103:H2 str. 12009	1	6.256250e-06	0.163	0.0965
E. coli O111:H- str. 11128	1	2.718269e-02	0.099	0.0931
E. coli O26:H11 str. 11368	1	4.070000e-07	0.1811	0.0943
E. coli IHE3034	1	2.388571e-05	0.1525	0.0968
E. coli ATCC 8739	1	7.538235e-06	0.1615	0.0963
E. coli 536	1	1.283333e-05	0.1578	0.0968
E. coli O157:H7 str. Sakai	1	6.128571e-06	0.1635	0.0963
E. coli S88	1	4.735366e-05	0.1481	0.0969
E. coli SE11	1	3.440000e-06	0.1678	0.0963
E. coli SE15	1	6.932558e-05	0.1345	0.0895
E. coli SMS-3-5	1	1.283333e-05	0.1553	0.0956
E. coli O157:H7 str. TW14359	1	7.700000e-06	0.1611	0.0965
E. coli UMN026	1	9.500000e-06	0.1569	0.095
E. coli W [iECW_1372]	1	2.151111e-06	0.1699	0.0959
E. coli KO11FL	1	6.128571e-06	0.1645	0.0966
E. coli ETEC H10407	1	7.791667e-07	0.1776	0.0949
E. coli O55:H7 str. CB9615	1	3.440000e-06	0.164	0.0943
E. coli K-12 MG1655 [iJO1366]	1	8.195000e-06	0.1473	0.0885
E. coli K-12 MG1655 [iJR904]	1	3.159804e-03	0.1209	0.0974
E. coli LF82	1	6.038776e-04	0.1174	0.0863
Mycobacterium tuberculosis H37Rv	0	3.025000e-01	0.097	0.123
E. coli O83:H1 str. NRG 857C	1	1.335172e-05	0.1573	0.0972
Shigella flexneri 2a str. 2457T	1	1.335172e-05	0.1534	0.0947
Staphylococcus aureus N315	1	7.403000e-04	0.1622	0.1206
Shigella flexneri 5 str. 8401	1	2.705405e-05	0.1515	0.0968
E. coli UM146	1	2.612500e-05	0.1521	0.0969
E. coli UMNK88	1	2.313235e-05	0.1531	0.0969
E. coli UTI89	1	1.283333e-05	0.1573	0.0968
E. coli W [iWFL 1372]	1	2.151111e-06	0.1699	0.0959
E. coli str. K-12 W3110	1	9.361000e-07	0.176	0.0954
Klebsiella pneumoniae_MGH78578	1	6.985000e-16	0.2322	0.0864
Bacillus subtilis str. 168	0	4.620000e-01	0.0729	0.1019
E. coli O157:H7 str. EDL933	1	1.987857e-06	0.1728	0.0965
Salmonella Typhimurium str. LT2	1	4.755208e-04	0.1392	0.1011

Table S5: Clustering of the conditionally essential genes. Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, whether the null hypothesis of uniform distribution of the conditionally essential genes is rejected by Kuiper's test (1) or not (0), the P-value of the test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 52 among the 55 genomes the null hypothesis is rejected, i.e., conditionally essential genes are significantly clustered. We consider a metabolic gene conditionally essential, if its deletion abolishes viability on at least one carbon source. Note that clustering of conditionally essential genes may be of limited biological relevance, because different conditionally essential genes may not be essential on the same set of carbon sources.

Species (strain)	Number of synthetic lethal gene pairs	Minimum genomic distance between synthetic lethal genes	Number of synthetic lethal pairs with genomic distance <50
E.coli K-12 MG1655 [iAF1260]	35	51	0
Methanosarcina barkeri str. Fusaro	71	0	14
Geobacter metallireducens GS-15	89	0	18
E. coli APEC O1	59	69	0
E. coli BL21(DE3) [iB21 1397]	40	51	0
E. coli BW2952	68	65	0
E. coli CFT073	191	0	3
E. coli O127:H6	59	69	0
E. coli 042	60	51	0
E. coli 55989	60	51	0
E. coli ABU 83972	59	69	0
E. coli B str. REL606	60	51	0
E. coli BL21-Gold(DE3)pLysS AG	40	50	0
E. coli BL21(DE3) [iECD1391]	40	51	0
E. coli DH1 [iEcDH1 1363]	60	53	0
E. coli DH1 [iECDH1ME8569 1439]	61	54	0
E. coli E24377A	59	51	0
E. coli ED1a	59	64	0
E. coli O157:H7	60	51	0
E. coli HS	60	51	0
E. coli NA114	59	0	2
E. coli O103:H2 str. 12009	61	0	3
E. coli O111:H- str. 11128	60	52	0
E. coli O26:H11 str. 11368	60	52	0
E. coli IHE3034	59	67	0
E. coli ATCC 8739	60	52	0
E. coli 536	59	55	0
E. coli O157:H7 str. Sakai	60	51	0
E. coli S88	59	68	0
E. coli SE11	60	51	0
E. coli SE15	67	67	0
E. coli SMS-3-5	60	52	0
E. coli O157:H7 str. TW14359	60	51	0
E. coli UMN026	56	51	0
E. coli W [iECW_1372]	61	0	2
E. coli KO11FL	61	52	0
E. coli ETEC H10407	60	51	0
E. coli O55:H7 str. CB9615	60	51	0
E. coli K-12 MG1655 [iJO1366]	70	54	0
E. coli K-12 MG1655 [iJR904]	20	4	4
E. coli LF82	59	0	19
Mycobacterium tuberculosis H37Rv	41	1	9
E. coli O83:H1 str. NRG 857C	59	68	0
Shigella flexneri 2a str. 2457T	62	63	0
Staphylococcus aureus N315	21	0	5
Shigella flexneri 5 str. 8401	60	67	0
E. coli UM146	60	57	0
E. coli UMNK88	60	0	2
E. coli UTI89	59	57	0
E. coli W [iWFL 1372]	61	0	2
E. coli str. K-12 W3110	60	54	0
Klebsiella pneumoniae_MGH78578	37	0	9
Bacillus subtilis str. 168	40	0	8
E. coli O157:H7 str. EDL933	60	51	0
Salmonella Typhimurium str. LT2	30	0	9

Table S6: Strictly synthetic lethal gene pairs. Each row corresponds to one of the 55 bacterial species or strains. Columns, from left to right, show the species name, the number of strictly synthetic lethal gene pairs, the distance between the synthetic lethal pairs with the shortest distance (smallest number of intervening genes) in the genome, and the number of strictly synthetic lethal gene pairs with distance below 50 intervening genes. In 41 of the genomes (82%), there are no strictly synthetic lethal gene pairs with distance below 50. We consider a pair of non-essential genes as strictly synthetic lethal if their simultaneous deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
E.coli K-12 MG1655 [iAF1260]	574538	48713	35	0	1.440476e-01	0	0
Methanosarcina barkeri str. Fusaro	130875	21682	57	14	2.059574e-01	1.4826	0
Geobacter metallireducens GS-15	232023	26009	71	18	3.058537e-02	2.2616	1
E. coli APEC O1	614115	49954	59	0	3.058537e-02	0	1
E. coli BL21(DE3) [iB21 1397]	636626	50712	40	0	1.442222e-01	0	0
E. coli BW2952	577882	46453	68	0	3.058537e-02	0	1
E. coli CFT073	561681	45881	188	3	9.001667e-03	0.1954	1
E. coli O127:H6	582591	48476	59	0	3.058537e-02	0	1
E. coli 042	616382	49993	60	0	3.058537e-02	0	1
E. coli 55989	634193	50782	60	0	3.058537e-02	0	1
E. coli ABU 83972	621948	50213	59	0	3.058537e-02	0	1
E. coli B str. REL606	633074	50731	60	0	3.058537e-02	0	1
E. coli BL21-Gold(DE3)pLysS AG	655898	51517	40	0	1.442222e-01	0	0
E. coli BL21(DE3) [iECD1391]	632135	50521	40	0	1.442222e-01	0	0
E. coli DH1 [iEcDH1 1363]	671776	52370	60	0	3.058537e-02	0	1
E. coli DH1 [iECDH1ME8569 1439]	763580	56199	61	0	3.058537e-02	0	1
E. coli E24377A	648766	51511	59	0	3.058537e-02	0	1
E. coli ED1a	576221	48123	59	0	3.058537e-02	0	1
E. coli O157:H7	560227	47466	60	0	3.058537e-02	0	1
E. coli HS	624135	50346	60	0	3.058537e-02	0	1
E. coli NA114	603008	49586	57	2	3.351887e-01	0.4267	0
E. coli O103:H2 str. 12009	630872	50595	58	3	6.260000e-01	0.645	0
E. coli O111:H- str. 11128	631969	50667	60	0	3.058537e-02	0	1
E. coli O26:H11 str. 11368	662610	51940	60	0	3.058537e-02	0	1
E. coli IHE3034	604285	49452	59	0	3.058537e-02	0	1
E. coli ATCC 8739	677564	52612	60	0	3.058537e-02	0	1
E. coli 536	609776	49691	59	0	3.058537e-02	0	1
E. coli O157:H7 str. Sakai	602108	49343	60	0	3.058537e-02	0	1
E. coli S88	605384	49497	59	0	3.058537e-02	0	1
E. coli SE11	654561	51645	60	0	3.058537e-02	0	1
E. coli SE15	576901	46318	67	0	3.058537e-02	0	1
E. coli SMS-3-5	653452	51566	60	0	3.058537e-02	0	1
E. coli O157:H7 str. TW14359	599925	49245	60	0	3.058537e-02	0	1
E. coli UMN026	637525	50970	56	0	3.058537e-02	0	1
E. coli W [iECW_1372]	683337	52893	59	2	3.351887e-01	0.4379	0
E. coli KO11FL	662537	52012	61	0	3.058537e-02	0	1
E. coli ETEC H10407	637566	50925	60	0	3.058537e-02	0	1
E. coli O55:H7 str. CB9615	582568	48498	60	0	3.058537e-02	0	1
E. coli K-12 MG1655 [iJO1366]	623648	48502	70	0	3.058537e-02	0	1
E. coli K-12 MG1655 [iJR904]	263757	32288	16	4	2.996939e-01	2.0422	0
E. coli LF82	579179	47402	40	19	1.969000e-06	5.8038	1
Mycobacterium tuberculosis H37Rv	91476	15899	32	9	2.165625e-01	1.6182	0
E. coli O83:H1 str. NRG 857C	611979	49787	59	0	3.058537e-02	0	1
Shigella flexneri 2a str. 2457T	477361	43292	57	0	3.058537e-02	0	1
Staphylococcus aureus N315	130683	24699	16	5	3.707407e-01	1.6534	0
Shigella flexneri 5 str. 8401	476324	43306	60	0	3.058537e-02	0	1
E. coli UM146	621883	50277	60	0	3.058537e-02	0	1
E. coli UMNK88	660439	51722	58	2	3.351887e-01	0.4403	0
E. coli UTI89	610793	49823	59	0	3.058537e-02	0	1
E. coli W [iWFL 1372]	683337	52893	59	2	3.351887e-01	0.4379	0
E. coli str. K-12 W3110	665915	52226	60	0	3.058537e-02	0	1
Klebsiella pneumoniae MGH78578	597269	51924	28	9	2.901250e-02	3.6973	1
Bacillus subtilis str. 168	231107	30579	32	8	1.578261e-01	1.8894	0
E. coli O157:H7 str. EDL933	609796	49670	60	0	3.058537e-02	0	1
Salmonella Typhimurium str. LT2	589406	49579	21	9	8.882500e-03	5.095	1

Table S7: Repulsion of the strictly synthetic lethal gene pairs. Each row corresponds to one of the 55 bacterial species or strains. Columns, from left to right, show the species (strain) name, the number of non-essential metabolic gene pairs that are neither (strictly) synthetic lethal nor less than 50 metabolic genes apart ($--$), the number of non-essential metabolic gene pairs that are not (strictly) synthetic lethal but less than 50 metabolic genes apart ($+-$), the number of non-essential metabolic gene pairs that are (strictly) synthetic lethal and less than 50 metabolic genes apart ($-+$), the number of non-essential metabolic gene pairs that are both (strictly) synthetic lethal and less than 50 metabolic genes apart ($++$), the P value of Fisher's exact test, the odds ratio (the odds of being synthetic lethal for pairs of non-essential metabolic genes with distance below 50, divided by the odds of being synthetic lethal for pairs of non-essential metabolic genes with distance below 50), and whether (strictly) synthetic lethal gene pairs are in significant repulsion. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). We consider a pair of non-essential genes as strictly synthetic lethal if their simultaneous deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species (strain)	Number of synthetic lethal gene pairs	Minimum genomic distance between synthetic lethal genes	Number of synthetic lethal pairs with genomic distance <50
E.coli K-12 MG1655 [iAF1260]	696	0	46
Methanosarcina barkeri str. Fusaro	71	0	14
Geobacter metallireducens GS-15	89	0	18
E. coli APEC O1	640	0	39
E. coli BL21(DE3) [iB21 1397]	614	0	41
E. coli BW2952	652	0	51
E. coli CFT073	191	0	3
E. coli O127:H6	720	0	52
E. coli 042	684	0	45
E. coli 55989	666	0	46
E. coli ABU 83972	660	0	38
E. coli B str. REL606	633	0	43
E. coli BL21-Gold(DE3)pLysS AG	614	0	41
E. coli BL21(DE3) [iECD1391]	624	0	53
E. coli DH1 [iEcDH1 1363]	626	0	40
E. coli DH1 [iECDH1ME8569 1439]	625	0	39
E. coli E24377A	640	0	55
E. coli ED1a	636	0	39
E. coli O157:H7	699	0	45
E. coli HS	697	0	44
E. coli NA114	647	0	34
E. coli O103:H2 str. 12009	633	0	53
E. coli O111:H- str. 11128	689	0	49
E. coli O26:H11 str. 11368	658	0	48
E. coli IHE3034	640	0	32
E. coli ATCC 8739	620	0	52
E. coli 536	658	0	40
E. coli O157:H7 str. Sakai	652	0	49
E. coli S88	657	0	33
E. coli SE11	642	0	49
E. coli SE15	662	0	39
E. coli SMS-3-5	650	0	47
E. coli O157:H7 str. TW14359	651	0	47
E. coli UMN026	686	0	47
E. coli W [iECW_1372]	644	0	49
E. coli KO11FL	652	0	51
E. coli ETEC H10407	632	0	44
E. coli O55:H7 str. CB9615	741	0	51
E. coli K-12 MG1655 [iJO1366]	642	0	47
E. coli K-12 MG1655 [iJR904]	348	0	23
E. coli LF82	671	0	78
Mycobacterium tuberculosis H37Rv	41	1	9
E. coli O83:H1 str. NRG 857C	670	0	39
Shigella flexneri 2a str. 2457T	713	0	58
Staphylococcus aureus N315	322	0	45
Shigella flexneri 5 str. 8401	626	0	44
E. coli UM146	633	0	32
E. coli UMNK88	622	0	37
E. coli UTI89	660	0	37
E. coli W [iWFL 1372]	644	0	49
E. coli str. K-12 W3110	641	0	43
Klebsiella pneumoniae MGH78578	576	0	37
Bacillus subtilis str. 168	552	0	51
E. coli O157:H7 str. EDL933	653	0	50
Salmonella Typhimurium str. LT2	226	0	42

Table S8: Conditionally synthetic lethal gene pairs. Each row corresponds to one of the 55 bacterial species or strains. Columns, from left to right, show the species (strain) name, the number of conditionally synthetic lethal gene pairs, the distance between the synthetic lethal pairs with the shortest distance (smallest number of intervening genes) in the genome, and the number of conditionally synthetic lethal gene pairs with distance below 50. In 54 genomes, conditionally synthetic lethal gene pairs exist whose member genes are adjacent in the genome. We consider a pair of non-essential genes as conditionally synthetic lethal if their simultaneous deletion abolishes viability on some but not all carbon sources.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
E.coli K-12 MG1655 [iAF1260]	573923	48667	650	46	5.578571e-01	0.8346	0
Methanosarcina barkeri str. Fusaro	130875	21682	57	14	5.342857e-01	1.4826	0
Geobacter metallireducens GS-15	232023	26009	71	18	5.445000e-02	2.2616	0
E. coli APEC O1	613572	49916	601	39	5.342857e-01	0.7977	0
E. coli BL21(DE3) [iB21 1397]	636092	50672	573	41	8.655952e-01	0.8982	0
E. coli BW2952	577348	46403	601	51	8.873333e-01	1.0558	0
E. coli CFT073	561681	45881	188	3	9.001667e-03	0.1954	1
E. coli O127:H6	581981	48425	668	52	8.873333e-01	0.9355	0
E. coli 042	615802	49949	639	45	7.522581e-01	0.8682	0
E. coli 55989	633632	50737	620	46	8.873333e-01	0.9266	0
E. coli ABU 83972	621384	50176	622	38	4.484615e-01	0.7566	0
E. coli B str. REL606	632543	50689	590	43	8.655952e-01	0.9095	0
E. coli BL21-Gold(DE3)pLysS AG	655364	51477	573	41	8.655952e-01	0.911	0
E. coli BL21(DE3) [iECD1391]	631603	50469	571	53	5.578571e-01	1.1616	0
E. coli DH1 [iEcDH1 1363]	671249	52331	586	40	8.116667e-01	0.8756	0
E. coli DH1 [iECDH1ME8569 1439]	763054	56161	586	39	8.655952e-01	0.9042	0
E. coli E24377A	648240	51456	585	55	5.578571e-01	1.1844	0
E. coli ED1a	575682	48085	597	39	5.342857e-01	0.7821	0
E. coli O157:H7	559632	47422	654	45	5.342857e-01	0.812	0
E. coli HS	623541	50303	653	44	5.578571e-01	0.8352	0
E. coli NA114	602452	49554	613	34	1.558333e-01	0.6743	0
E. coli O103:H2 str. 12009	630350	50545	580	53	6.865517e-01	1.1396	0
E. coli O111:H- str. 11128	631388	50619	640	49	9.282653e-01	0.955	0
E. coli O26:H11 str. 11368	662059	51893	610	48	9.816981e-01	1.0039	0
E. coli IHE3034	603735	49421	608	32	1.045000e-01	0.643	0
E. coli ATCC 8739	677055	52561	568	52	5.578571e-01	1.1793	0
E. coli 536	609216	49652	618	40	5.342857e-01	0.7942	0
E. coli O157:H7 str. Sakai	601564	49295	603	49	1	0.9916	0
E. coli S88	604818	49465	624	33	1.045000e-01	0.6466	0
E. coli SE11	654027	51597	593	49	8.905319e-01	1.0474	0
E. coli SE15	576344	46280	623	39	5.342857e-01	0.7796	0
E. coli SMS-3-5	652908	51520	603	47	1	0.9878	0
E. coli O157:H7 str. TW14359	599380	49199	604	47	9.282653e-01	0.948	0
E. coli UMN026	636942	50923	639	47	8.655952e-01	0.92	0
E. coli W [iECW_1372]	682801	52846	595	49	8.655952e-01	1.064	0
E. coli KO11FL	661996	51962	601	51	8.655952e-01	1.0811	0
E. coli ETEC H10407	637037	50882	588	44	8.905319e-01	0.9369	0
E. coli O55:H7 str. CB9615	581937	48448	690	51	7.700000e-01	0.8878	0
E. coli K-12 MG1655 [iJO1366]	623123	48455	595	47	9.511765e-01	1.0158	0
E. coli K-12 MG1655 [iJR904]	263448	32269	325	23	1.045000e-01	0.5778	0
E. coli LF82	578626	47343	593	78	5.032500e-03	1.6076	1
Mycobacterium tuberculosis H37Rv	91476	15899	32	9	5.342857e-01	1.6182	0
E. coli O83:H1 str. NRG 857C	611406	49749	631	39	4.484615e-01	0.7596	0
Shigella flexneri 2a str. 2457T	476763	43234	655	58	9.816981e-01	0.9765	0
Staphylococcus aureus N315	130422	24659	277	45	7.351667e-01	0.8592	0
Shigella flexneri 5 str. 8401	475801	43263	582	44	5.578571e-01	0.8315	0
E. coli UM146	621341	50246	601	32	1.299375e-01	0.6584	0
E. coli UMNK88	659912	51687	585	37	5.578571e-01	0.8075	0
E. coli UTI89	610228	49787	623	37	3.558500e-01	0.7279	0
E. coli W [iWFL 1372]	682801	52846	595	49	8.655952e-01	1.064	0
E. coli str. K-12 W3110	665376	52184	598	43	8.655952e-01	0.9168	0
Klebsiella pneumoniae MGH78578	596758	51896	539	37	5.342857e-01	0.7894	0
Bacillus subtilis str. 168	230638	30536	501	51	3.675000e-01	0.7689	0
E. coli O157:H7 str. EDL933	609252	49621	603	50	9.511765e-01	1.0181	0
Salmonella Typhimurium str. LT2	589243	49546	184	42	7.040000e-06	2.7147	1

Table S9: Repulsion of conditionally synthetic lethal gene pairs. Each row corresponds to one of the 55 bacterial species or strains. Columns, from left to right, show the species (strain) name, the number of non-essential metabolic gene pairs that are neither (conditionally) synthetic lethal nor less than 50 metabolic genes apart ($--$), the number of non-essential metabolic gene pairs that are not (conditionally) synthetic lethal but less than 50 metabolic genes apart ($+-$), the number of non-essential metabolic gene pairs that are (conditionally) synthetic lethal and less than 50 metabolic genes apart ($-+$), the number of non-essential metabolic gene pairs that are both (conditionally) synthetic lethal and less than 50 metabolic genes apart ($++$), the P value of Fisher's exact test, the odds ratio (the odds of being (conditionally) synthetic lethal among the pairs of non-essential metabolic genes with less than 50 metabolic genes apart divided by the odds of being (conditionally) synthetic lethal among the pairs of non-essential metabolic genes with more than or equal to 50 metabolic genes apart), and whether (conditionally) synthetic lethal gene pairs are in significant repulsion. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). We consider a pair of non-essential genes as conditionally synthetic lethal if their simultaneous deletion abolishes viability on some but not all carbon sources.

Species (strain)	Number of clusters	Average length	Max length	No. Strictly non-essential	Fraction strictly non-essential	No. Conditionally non-essential	Fraction conditionally non-essential
E.coli K-12 MG1655 [iAF1260]	148	6.38	48	147	0.99	148	1
Methanosarcina barkeri str. Fusaro	81	6.89	33	78	0.96	78	0.96
Geobacter metallireducens GS-15	103	7.09	92	98	0.95	98	0.95
E. coli APEC O1	166	6.04	47	163	0.98	165	0.99
E. coli BL21(DE3) [iB21 1397]	154	6.62	47	151	0.98	153	0.99
E. coli BW2952	170	5.72	30	167	0.98	169	0.99
E. coli CFT073	107	10.4	70	104	0.97	104	0.97
E. coli O127:H6	153	6.41	45	150	0.98	152	0.99
E. coli 042	153	6.51	45	150	0.98	152	0.99
E. coli 55989	157	6.45	66	155	0.99	156	0.99
E. coli ABU 83972	155	6.52	47	152	0.98	154	0.99
E. coli B str. REL606	154	6.57	47	151	0.98	153	0.99
E. coli BL21-Gold(DE3)pLysS AG	154	6.73	47	150	0.97	152	0.99
E. coli BL21(DE3) [iECD1391]	155	6.55	47	152	0.98	154	0.99
E. coli DH1 [iEcDH1 1363]	151	6.95	47	147	0.97	149	0.99
E. coli DH1 [iECDH1ME8569 1439]	150	7.52	97	147	0.98	149	0.99
E. coli E24377A	154	6.68	66	150	0.97	154	1
E. coli ED1a	156	6.21	47	153	0.98	155	0.99
E. coli O157:H7	152	6.24	44	151	0.99	151	0.99
E. coli HS	154	6.51	47	152	0.99	153	0.99
E. coli NA114	155	6.39	47	152	0.98	154	0.99
E. coli O103:H2 str. 12009	154	6.62	66	152	0.99	153	0.99
E. coli O111:H- str. 11128	154	6.62	65	150	0.97	153	0.99
E. coli O26:H11 str. 11368	156	6.67	66	154	0.99	155	0.99
E. coli IHE3034	157	6.33	47	154	0.98	156	0.99
E. coli ATCC 8739	151	6.98	47	147	0.97	149	0.99
E. coli 536	158	6.32	47	155	0.98	157	0.99
E. coli O157:H7 str. Sakai	154	6.45	66	152	0.99	153	0.99
E. coli S88	154	6.47	46	151	0.98	153	0.99
E. coli SE11	153	6.76	67	150	0.98	152	0.99
E. coli SE15	178	5.4	36	175	0.98	177	0.99
E. coli SMS-3-5	154	6.68	47	151	0.98	153	0.99
E. coli O157:H7 str. TW14359	154	6.42	66	153	0.99	153	0.99
E. coli UMN026	156	6.47	47	153	0.98	156	1
E. coli W [iECW_1372]	152	6.98	62	150	0.99	151	0.99
E. coli KO11FL	153	6.81	66	150	0.98	151	0.99
E. coli ETEC H10407	152	6.73	47	149	0.98	151	0.99
E. coli O55:H7 str. CB9615	154	6.24	44	152	0.99	153	0.99
E. coli K-12 MG1655 [iJO1366]	175	5.69	32	174	0.99	175	1
E. coli K-12 MG1655 [iJR904]	135	4.39	29	134	0.99	135	1
E. coli LF82	129	7.71	60	111	0.86	121	0.94
Mycobacterium tuberculosis H37Rv	104	4.46	22	104	1	104	1
E. coli O83:H1 str. NRG 857C	152	6.61	47	149	0.98	151	0.99
Shigella flexneri 2a str. 2457T	148	6.04	29	145	0.98	146	0.99
Staphylococcus aureus N315	82	5.59	21	80	0.98	81	0.99
Shigella flexneri 5 str. 8401	149	6.03	40	147	0.99	148	0.99
E. coli UM146	155	6.52	47	152	0.98	154	0.99
E. coli UMNK88	168	6.21	47	165	0.98	167	0.99
E. coli UTI89	157	6.37	47	154	0.98	156	0.99
E. coli W [iWFL 1372]	152	6.98	62	150	0.99	151	0.99
E. coli str. K-12 W3110	155	6.73	47	153	0.99	154	0.99
Klebsiella pneumoniae MGH78578	121	8.1	39	116	0.96	118	0.98
Bacillus subtilis str. 168	108	5.38	34	104	0.96	107	0.99
E. coli O157:H7 str. EDL933	151	6.62	67	149	0.99	150	0.99
Salmonella Typhimurium str. LT2	139	7.24	61	135	0.97	138	0.99

Table S10: Non-essential clusters of strictly non-essential genes. Each row corresponds to one of the 55 bacterial species or strains. Columns, show the species (strain) name (first column), the number of the clusters of strictly non-essential genes (second column), the average length of the clusters of strictly non-essential genes (third column), the length of the largest cluster of non-essential genes (fourth column), the number (fifth column) and fraction (sixth column) of strictly non-essential clusters of strictly non-essential genes, and the number (seventh column) and fraction (eighth column) of conditionally non-essential clusters of strictly non-essential genes. We consider a metabolic gene as conditionally non-essential if its deletion does not abolish viability on at least one carbon source, and consider a metabolic gene as strictly non-essential if its deletion does not abolish viability on any carbon source. A *cluster of conditionally non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that lies between two nearest non-adjacent strictly essential metabolic genes. Likewise, a *cluster of strictly non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that lies between two nearest non-adjacent conditionally essential metabolic genes. Finally, we consider a cluster of strictly non-essential genes as a *strictly non-essential cluster of strictly non-essential genes* if simultaneous deletion of all the genes in the cluster does not abolish viability on any carbon source, and we consider a cluster of strictly non-essential genes as a*conditionally non-essential cluster of strictly non-essential genes* if simultaneous deletion of all the genes in the cluster does not abolish viability on at least one carbon source.

Species (strain)	Number of clusters	Average length	Max length	No. Strictly non-essential	Fraction strictly non-essential	No. Conditionally non-essential	Fraction conditionally non-essential
E.coli K-12 MG1655 [iAF1260]	81	13.79	91	39	0.48	76	0.94
Methanosarcina barkeri str. Fusaro	81	6.89	33	78	0.96	78	0.96
Geobacter metallireducens GS-15	103	7.09	92	98	0.95	98	0.95
E. coli APEC O1	94	12.28	73	48	0.51	88	0.94
E. coli BL21(DE3) [iB21 1397]	87	13.49	74	47	0.54	81	0.93
E. coli BW2952	109	10.28	63	67	0.61	103	0.94
E. coli CFT073	107	10.4	70	104	0.97	104	0.97
E. coli O127:H6	86	13.08	76	47	0.55	78	0.91
E. coli 042	86	13.44	71	46	0.53	79	0.92
E. coli 55989	86	13.63	70	46	0.53	79	0.92
E. coli ABU 83972	86	13.5	72	46	0.53	78	0.91
E. coli B str. REL606	86	13.62	71	45	0.52	79	0.92
E. coli BL21-Gold(DE3)pLysS AG	87	13.69	73	47	0.54	81	0.93
E. coli BL21(DE3) [iECD1391]	87	13.45	73	47	0.54	81	0.93
E. coli DH1 [iEcDH1 1363]	85	14.18	80	44	0.52	79	0.93
E. coli DH1 [iECDH1ME8569 1439]	84	15.27	98	43	0.51	78	0.93
E. coli E24377A	86	13.76	74	46	0.53	80	0.93
E. coli ED1a	87	12.86	70	47	0.54	80	0.92
E. coli O157:H7	86	12.84	67	45	0.52	80	0.93
E. coli HS	86	13.52	74	46	0.53	79	0.92
E. coli NA114	85	13.46	86	48	0.56	77	0.91
E. coli O103:H2 str. 12009	86	13.59	66	48	0.56	79	0.92
E. coli O111:H- str. 11128	86	13.6	91	48	0.56	79	0.92
E. coli O26:H11 str. 11368	86	13.92	66	46	0.53	80	0.93
E. coli IHE3034	86	13.31	73	46	0.53	78	0.91
E. coli ATCC 8739	85	14.24	75	44	0.52	79	0.93
E. coli 536	87	13.22	72	47	0.54	79	0.91
E. coli O157:H7 str. Sakai	87	13.14	66	46	0.53	81	0.93
E. coli S88	86	13.33	72	46	0.53	78	0.91
E. coli SE11	85	14	70	45	0.53	78	0.92
E. coli SE15	109	10.26	63	67	0.61	102	0.94
E. coli SMS-3-5	86	13.83	74	45	0.52	79	0.92
E. coli O157:H7 str. TW14359	87	13.11	66	46	0.53	81	0.93
E. coli UMN026	87	13.48	71	45	0.52	82	0.94
E. coli W [iECW_1372]	85	14.29	73	46	0.54	79	0.93
E. coli KO11FL	86	13.92	74	46	0.53	79	0.92
E. coli ETEC H10407	86	13.66	72	46	0.53	79	0.92
E. coli O55:H7 str. CB9615	86	13.08	67	44	0.51	79	0.92
E. coli K-12 MG1655 [iJO1366]	110	10.55	65	65	0.59	106	0.96
E. coli K-12 MG1655 [iJR904]	71	10.86	52	28	0.39	68	0.96
E. coli LF82	70	16.33	86	33	0.47	57	0.81
Mycobacterium tuberculosis H37Rv	104	4.46	22	104	1	104	1
E. coli O83:H1 str. NRG 857C	86	13.4	72	46	0.53	78	0.91
Shigella flexneri 2a str. 2457T	89	11.51	70	50	0.56	83	0.93
Staphylococcus aureus N315	36	15.58	64	15	0.42	29	0.81
Shigella flexneri 5 str. 8401	88	11.6	68	51	0.58	82	0.93
E. coli UM146	86	13.5	71	47	0.55	78	0.91
E. coli UMNK88	105	11.38	76	64	0.61	98	0.93
E. coli UTI89	87	13.23	74	47	0.54	79	0.91
E. coli W [iWFL 1372]	85	14.29	73	46	0.54	79	0.93
E. coli str. K-12 W3110	86	13.95	77	45	0.52	79	0.92
Klebsiella pneumoniae MGH78578	54	21.24	127	21	0.39	48	0.89
Bacillus subtilis str. 168	52	13.94	72	25	0.48	46	0.88
E. coli O157:H7 str. EDL933	86	13.37	67	46	0.53	80	0.93
Salmonella Typhimurium str. LT2	77	14.71	99	41	0.53	71	0.92

Table S11: Non-essential clusters of conditionally non-essential genes. Each row corresponds to one of the 55 bacterial species or strains. Columns, show the species (strain) name (first column), the number of the clusters of conditionally non-essential genes (second column), the average length of the clusters of conditionally non-essential genes (third column), the length of the largest cluster of non-essential genes (fourth column), the number (fifth column) and fraction (sixth column) of strictly non-essential clusters of conditionally non-essential genes, and the number (seventh column) and fraction (eighth column) of conditionally non-essential clusters of conditionally non-essential genes. We consider a metabolic gene as conditionally non-essential if its deletion does not abolish viability on at least one carbon source, and consider a metabolic gene as strictly non-essential if its deletion does not abolish viability on any carbon source. A *cluster of conditionally non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that lies between two nearest non-adjacent strictly essential metabolic genes. Likewise, a *cluster of strictly non-essential metabolic genes* is a set of consecutive non-essential metabolic genes that lies between two nearest non-adjacent conditionally essential metabolic genes. Finally, we consider a cluster of conditionally non-essential genes as *strictly non-essential cluster of conditionally non-essential genes* if simultaneous deletion of all the genes in the cluster does not abolish viability on any carbon sources, and we consider a cluster of conditionally non-essential genes as *conditionally non-essential cluster of conditionally non-essential genes* if simultaneous deletion of all the genes in the cluster does not abolish viability on at least one carbon source.

Species (strains)	Number of operons	Average number of genes per operon	Number of operons with at least one conditionally essential gene	Fraction of operons with at least one conditionally essential gene	Number of operons with more than one conditionally essential gene	Fraction of operons with more than one conditionally essential gene	Number of operons with at least one strictly essential gene	Fraction of operons with at least one strictly essential gene	Number of operons with more than one strictly essential gene	Fraction of operons with more than one strictly essential gene
E.coli K-12 MG1655 [iAF1260]	241	3.07	95	0.39	62	0.26	48	0.2	25	0.1
Methanosarcina barkeri str. Fusaro	91	2.75	20	0.22	15	0.16	20	0.22	15	0.16
Geobacter metallireducens GS-15	203	3.12	67	0.33	47	0.23	67	0.33	47	0.23
E. coli APEC O1	247	2.94	101	0.41	59	0.24	52	0.21	29	0.12
E. coli BL21(DE3) [iB21 1397]	277	3.19	108	0.39	63	0.23	59	0.21	31	0.11
E. coli BW2952	247	3.14	113	0.46	71	0.29	69	0.28	41	0.17
E. coli CFT073	260	2.8	61	0.23	31	0.12	61	0.23	31	0.12
E. coli O127:H6	244	3.01	97	0.4	59	0.24	53	0.22	29	0.12
E. coli 042	252	3.07	100	0.4	64	0.25	54	0.21	30	0.12
E. coli 55989	257	3.05	101	0.39	61	0.24	55	0.21	29	0.11
E. coli ABU 83972	247	3.07	99	0.4	61	0.25	51	0.21	29	0.12
E. coli B str. REL606	256	3	99	0.39	62	0.24	55	0.21	29	0.11
E. coli BL21-Gold(DE3)pLysS AG	256	3.05	98	0.38	62	0.24	56	0.22	31	0.12
E. coli BL21(DE3) [iECD1391]	253	3.02	99	0.39	63	0.25	56	0.22	31	0.12
E. coli DH1 [iEcDH1 1363]	260	3.09	101	0.39	60	0.23	56	0.22	29	0.11
E. coli DH1 [iECDH1ME8569 1439]	255	3.1	101	0.4	60	0.24	56	0.22	29	0.11
E. coli E24377A	260	2.98	99	0.38	61	0.23	54	0.21	29	0.11
E. coli ED1a	242	3	103	0.43	57	0.24	53	0.22	28	0.12
E. coli O157:H7	240	2.95	98	0.41	58	0.24	54	0.23	28	0.12
E. coli HS	253	3.06	100	0.4	62	0.25	55	0.22	29	0.11
E. coli NA114	245	3.03	98	0.4	60	0.24	52	0.21	29	0.12
E. coli O103:H2 str. 12009	250	3.06	99	0.4	59	0.24	55	0.22	29	0.12
E. coli O111:H- str. 11128	251	3.05	97	0.39	57	0.23	55	0.22	29	0.12
E. coli O26:H11 str. 11368	261	3.04	100	0.38	58	0.22	55	0.21	29	0.11
E. coli IHE3034	242	3.01	99	0.41	59	0.24	51	0.21	28	0.12
E. coli ATCC 8739	258	3.1	98	0.38	61	0.24	54	0.21	29	0.11
E. coli 536	248	2.97	100	0.4	61	0.25	52	0.21	29	0.12
E. coli O157:H7 str. Sakai	248	2.99	98	0.4	58	0.23	54	0.22	28	0.11
E. coli S88	248	3.01	99	0.4	60	0.24	51	0.21	29	0.12
E. coli SE11	259	3.08	100	0.39	60	0.23	55	0.21	29	0.11

E. coli SE15	251	3.07	117	0.47	73	0.29	67	0.27	41	0.16
E. coli SMS-3-5	260	3.08	103	0.4	66	0.25	55	0.21	30	0.12
E. coli O157:H7 str. TW14359	245	3	99	0.4	59	0.24	55	0.22	29	0.12
E. coli UMN026	259	3.05	100	0.39	64	0.25	54	0.21	30	0.12
E. coli W [iECW_1372]	259	3.03	100	0.39	61	0.24	56	0.22	29	0.11
E. coli KO11FL	258	3.09	98	0.38	61	0.24	54	0.21	29	0.11
E. coli ETEC H10407	257	3.01	100	0.39	60	0.23	55	0.21	29	0.11
E. coli O55:H7 str. CB9615	247	3.02	100	0.4	61	0.25	54	0.22	28	0.11
E. coli K-12 MG1655 [iJO1366]	262	3.12	115	0.44	75	0.29	68	0.26	41	0.16
E. coli K-12 MG1655 [iJR904]	181	2.87	81	0.45	58	0.32	35	0.19	21	0.12
E. coli LF82	246	3.02	98	0.4	60	0.24	53	0.22	28	0.11
Mycobacterium tuberculosis H37Rv	122	2.82	55	0.45	30	0.25	55	0.45	30	0.25
E. coli O83:H1 str. NRG 857C	248	3.03	99	0.4	61	0.25	53	0.21	29	0.12
Shigella flexneri 2a str. 2457T	218	2.93	90	0.41	55	0.25	57	0.26	31	0.14
Shigella flexneri 5 str. 8401	218	2.94	86	0.39	52	0.24	54	0.25	30	0.14
E. coli UM146	249	3	99	0.4	61	0.24	51	0.2	29	0.12
E. coli UMNK88	259	3.03	98	0.38	60	0.23	55	0.21	29	0.11
E. coli UTI89	259	2.95	101	0.39	60	0.23	53	0.2	29	0.11
Klebsiella pneumoniae_MGH78578	227	3	70	0.31	44	0.19	24	0.11	10	0.04
Bacillus subtilis str. 168	161	3.24	76	0.47	52	0.32	33	0.2	25	0.16
E. coli O157:H7 str. EDL933	253	3.02	98	0.39	59	0.23	54	0.21	28	0.11
Salmonella Typhimurium str. LT2	247	2.95	82	0.33	48	0.19	43	0.17	23	0.09

Table S12: Metabolic genes in Operons. Each row corresponds to a given species (strain). Columns, from left to right, show species or strain names, the total number of operons in the genome, the average number of genes per operon, the number of operons with at least one conditionally essential gene, the fraction of operons with at least one conditionally essential gene, the number of operons with more than one conditionally essential metabolic gene, the fraction of operons with more than one conditionally essential metabolic gene, the number of operons with at least one strictly essential gene, the fraction of operons with at least one strictly essential gene, the number of operons with more than one strictly essential metabolic gene, and the fraction of operons with more than one strictly essential metabolic gene.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
E.coli K-12 MG1655 [iAF1260]	470	647	51	93	1.659574e-01	1.3247	0
Methanosarcina barkeri str. Fusaro	348	211	94	39	8.646512e-02	0.6843	0
Geobacter metallireducens GS-15	271	459	83	174	1.885000e-01	1.2377	0
E. coli APEC O1	528	627	60	98	8.662727e-02	1.3754	0
E. coli BL21(DE3) [iB21 1397]	415	760	39	123	1.388706e-02	1.7222	1
E. coli BW2952	490	630	63	145	5.168800e-03	1.7901	1
E. coli CFT073	498	616	80	113	4.595102e-01	1.1419	0
E. coli O127:H6	497	629	52	106	1.534963e-02	1.6107	1
E. coli 042	488	669	52	105	3.825714e-02	1.4729	1
E. coli 55989	497	676	49	108	1.534963e-02	1.6205	1
E. coli ABU 83972	508	654	53	105	2.230526e-02	1.5389	1
E. coli B str. REL606	511	661	49	108	1.364000e-02	1.7039	1
E. coli BL21-Gold(DE3)pLysS AG	525	667	49	113	6.767429e-03	1.8152	1
E. coli BL21(DE3) [iECD1391]	518	653	50	112	7.222222e-03	1.7769	1
E. coli DH1 [iEcDH1 1363]	511	695	49	108	1.534963e-02	1.6206	1
E. coli DH1 [iECDH1ME8569 1439]	602	681	46	110	1.981200e-03	2.1139	1
E. coli E24377A	516	668	50	107	1.430000e-02	1.6531	1
E. coli ED1a	499	621	55	104	2.760000e-02	1.5194	1
E. coli O157:H7	502	603	51	106	1.224364e-02	1.7303	1
E. coli HS	498	666	49	108	1.430000e-02	1.6481	1
E. coli NA114	504	641	54	102	3.825714e-02	1.4852	1
E. coli O103:H2 str. 12009	512	658	50	107	1.388706e-02	1.6652	1
E. coli O111:H- str. 11128	513	658	49	108	1.224364e-02	1.7184	1
E. coli O26:H11 str. 11368	512	686	49	108	1.430000e-02	1.645	1
E. coli IHE3034	521	625	54	104	1.534963e-02	1.6055	1
E. coli ATCC 8739	517	694	50	107	1.623030e-02	1.5942	1
E. coli 536	518	633	54	104	1.623030e-02	1.576	1
E. coli O157:H7 str. Sakai	509	635	51	106	1.388706e-02	1.666	1
E. coli S88	506	641	53	105	1.901714e-02	1.5639	1
E. coli SE11	502	689	49	108	1.623030e-02	1.6059	1
E. coli SE15	492	627	64	144	5.168800e-03	1.7656	1
E. coli SMS-3-5	497	693	48	109	1.534963e-02	1.6286	1
E. coli O157:H7 str. TW14359	513	629	52	105	1.430000e-02	1.6468	1
E. coli UMN026	491	683	50	108	2.192432e-02	1.5528	1
E. coli W [iECW_1372]	539	677	47	109	6.621333e-03	1.8464	1
E. coli KO11FL	509	689	49	107	1.623030e-02	1.6132	1
E. coli ETEC H10407	511	665	49	108	1.364000e-02	1.6937	1
E. coli O55:H7 str. CB9615	485	641	51	106	1.865882e-02	1.5726	1
E. coli K-12 MG1655 [iJO1366]	487	673	62	145	7.222222e-03	1.6924	1
E. coli K-12 MG1655 [iJR904]	329	443	55	77	8.666667e-01	1.0397	0
E. coli LF82	507	637	53	105	1.623030e-02	1.5768	1
Mycobacterium tuberculosis H37Rv	233	231	84	113	1.023822e-01	1.3569	0
E. coli O83:H1 str. NRG 857C	507	646	52	106	1.623030e-02	1.5998	1
Shigella flexneri 2a str. 2457T	495	530	54	109	5.168800e-03	1.8852	1
Shigella flexneri 5 str. 8401	490	532	54	108	5.168800e-03	1.8421	1
E. coli UM146	519	643	52	105	1.430000e-02	1.6298	1
E. coli UMNK88	518	678	49	108	1.388706e-02	1.6839	1
E. coli UTI89	494	658	53	105	3.328000e-02	1.4874	1
Klebsiella pneumoniae MGH78578	514	633	35	47	7.612800e-01	1.0904	0
Bacillus subtilis str. 168	285	441	37	81	1.164348e-01	1.4148	0
E. coli O157:H7 str. EDL933	494	657	51	106	2.192432e-02	1.5628	1
Salmonella Typhimurium str. LT2	483	650	59	79	1	0.995	0

Table S13: Operons and the strictly essential genes. Each row corresponds to a given species or strain. Columns, from left to right, show species (strain) names, the number of metabolic genes that are neither strictly essential nor belong to an operon $(--)$, the number of metabolic genes that are not strictly essential but do belong to an operon $(-+)$, the number of metabolic genes that are strictly essential, but do not belong to an operon $(+-)$, the number of metabolic genes that are both strictly essential and belong to an operon $(++)$, the P value of a Fisher's exact test on this data, the odds ratio (defined as the odds of being strictly essential for operonic metabolic genes divided by the odds of being strictly essential for non-operonic metabolic genes), and whether the null hypothesis of a lack of association between a gene's strict essentiality and being part of an operon is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 42 of the 52 species (80.76%) the null hypothesis is rejected. Note that we consider a metabolic gene strictly essential, if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Strain (species)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
E.coli K-12 MG1655 [iAF1260]	411	533	110	207	6.552000e-03	1.4511	1
Methanosarcina barkeri str. Fusaro	348	211	94	39	7.910638e-02	0.6843	0
Geobacter metallireducens GS-15	271	459	83	174	1.809600e-01	1.2377	0
E. coli APEC O1	474	530	114	195	1.961818e-03	1.5298	1
E. coli BL21(DE3) [iB21 1397]	377	644	77	239	1.388000e-04	1.817	1
E. coli BW2952	437	536	116	239	1.846000e-04	1.6798	1
E. coli CFT073	498	616	80	113	4.414902e-01	1.1419	0
E. coli O127:H6	447	535	102	200	5.232500e-04	1.6383	1
E. coli 042	439	558	101	216	2.995200e-04	1.6825	1
E. coli 55989	442	571	104	213	9.017895e-04	1.5854	1
E. coli ABU 83972	459	552	102	207	2.995200e-04	1.6875	1
E. coli B str. REL606	456	557	104	212	3.000000e-04	1.6688	1
E. coli BL21-Gold(DE3)pLysS AG	471	567	103	213	1.846000e-04	1.7178	1
E. coli BL21(DE3) [iECD1391]	465	551	103	214	1.388000e-04	1.7534	1
E. coli DH1 [iEcDH1 1363]	456	594	104	209	1.646667e-03	1.5427	1
E. coli DH1 [iECDH1ME8569 1439]	548	580	100	211	1.216800e-05	1.9936	1
E. coli E24377A	464	566	102	209	2.995200e-04	1.6798	1
E. coli ED1a	453	516	101	209	1.326000e-04	1.8167	1
E. coli O157:H7	449	500	104	209	1.326000e-04	1.8046	1
E. coli HS	445	559	102	215	2.995200e-04	1.678	1
E. coli NA114	451	540	107	203	8.825946e-04	1.5845	1
E. coli O103:H2 str. 12009	461	559	101	206	2.995200e-04	1.682	1
E. coli O111:H- str. 11128	461	560	101	206	2.995200e-04	1.679	1
E. coli O26:H11 str. 11368	456	586	105	208	1.646667e-03	1.5415	1
E. coli IHE3034	468	527	107	202	2.995200e-04	1.6765	1
E. coli ATCC 8739	463	592	104	209	1.129333e-03	1.5717	1
E. coli 536	469	531	103	206	1.341600e-04	1.7665	1
E. coli O157:H7 str. Sakai	456	538	104	203	4.030000e-04	1.6544	1
E. coli S88	456	541	103	205	2.995200e-04	1.6776	1
E. coli SE11	450	586	101	211	7.473143e-04	1.6043	1
E. coli SE15	442	520	114	251	2.886000e-05	1.8715	1
E. coli SMS-3-5	449	580	96	222	1.341600e-04	1.7902	1
E. coli O157:H7 str. TW14359	459	531	106	203	3.447407e-04	1.6554	1
E. coli UMN026	438	573	103	218	5.247273e-04	1.6179	1
E. coli W [iECW_1372]	487	575	99	211	1.326000e-04	1.8051	1
E. coli KO11FL	456	586	102	210	7.473143e-04	1.6021	1
E. coli ETEC H10407	462	562	98	211	1.341600e-04	1.77	1
E. coli O55:H7 str. CB9615	435	527	101	220	1.326000e-04	1.798	1
E. coli K-12 MG1655 [iJO1366]	434	562	115	256	1.341600e-04	1.7191	1
E. coli K-12 MG1655 [iJR904]	262	331	122	189	1.666122e-01	1.2262	0
E. coli LF82	459	537	101	205	1.846000e-04	1.7349	1
Mycobacterium tuberculosis H37Rv	233	231	84	113	9.598333e-02	1.3569	0
E. coli O83:H1 str. NRG 857C	458	547	101	205	2.945647e-04	1.6995	1
Shigella flexneri 2a str. 2457T	441	454	108	185	4.344516e-04	1.6639	1
Shigella flexneri 5 str. 8401	437	462	107	178	1.391000e-03	1.5735	1
E. coli UM146	469	542	102	206	1.388000e-04	1.7476	1
E. coli UMNK88	466	579	101	207	4.073333e-04	1.6495	1
E. coli UTI89	442	559	105	204	1.850233e-03	1.5362	1
Klebsiella pneumoniae MGH78578	452	528	97	152	5.166087e-02	1.3415	0
Bacillus subtilis str. 168	246	336	76	186	4.073333e-04	1.7918	1
E. coli O157:H7 str. EDL933	443	558	102	205	8.377778e-04	1.5956	1
Salmonella Typhimurium str. LT2	432	575	110	154	7.270000e-01	1.0518	0

Table S14: Operons and the conditionally essential genes. Each row corresponds to a given species or strain. Columns, from left to right, show species (strain) names, the number of metabolic genes that are neither conditionally essential nor belong to an operon $(--)$, the number of metabolic genes that are not conditionally essential but do belong to an operon $(-+)$, the number of metabolic genes that are conditionally essential, but do not belong to an operon $(+-)$, the number of metabolic genes that are both conditionally essential and belong to an operon $(++)$, the P value of a Fisher exact test on this data, the odds ratio (defined as the odds of being conditionally essential for operonic metabolic genes divided by the odds of being conditionally essential for non-operonic metabolic genes), and whether the null hypothesis of a lack of association between a gene's conditional essentiality and being on an operon is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 45 of the 52 species (86.54%) the null hypothesis is rejected. Note that we consider a metabolic gene conditionally essential, if its deletion abolishes viability on at least one carbon source.

	Essential genes belonging to an operon				Essential genes not belonging to an operon				All essential genes			
Species(strains)	hypothesis rejected	P-value	Kuiper test statistic	Critical value	hypothesis rejected	P-value	Kuiper test statistic	Critical value	hypothesis rejected	P-value	Kuiper test statistic	Critical value
E.coli K-12 MG1655 [iAF1260]	1	2.015000e-03	0.232	0.1777	0	2.050286e-01	0.2121	0.2385	1	6.626977e-03	0.1713	0.1435
Methanosarcina barkeri str. Fusaro	1	3.650000e-02	0.2797	0.271	0	4.739091e-01	0.1316	0.1769	0	2.579608e-01	0.1211	0.1492
Geobacter metallireducens GS-15	1	3.605333e-05	0.22	0.1307	0	5.148000e-02	0.248	0.1876	1	4.784000e-04	0.1688	0.1078
E. coli APEC O1	1	1.495000e-05	0.303	0.1736	0	6.405957e-01	0.1476	0.22	1	7.210667e-04	0.2082	0.1371
E. coli BL21(DE3) [iB21 1397]	1	2.610196e-02	0.165	0.1551	0	2.704000e-01	0.2274	0.271	1	4.768780e-03	0.1656	0.1354
E. coli BW2952	1	2.696571e-04	0.2183	0.143	0	6.405957e-01	0.1442	0.2147	1	2.016182e-03	0.1694	0.1197
E. coli CFT073	1	1.035273e-03	0.222	0.1617	0	3.417143e-01	0.1529	0.191	1	2.016182e-03	0.1687	0.1242
E. coli O127:H6	1	2.746667e-03	0.2113	0.1668	0	1.686061e-01	0.2311	0.2362	1	3.105556e-03	0.1733	0.1371
E. coli 042	1	2.015000e-03	0.2184	0.1676	0	1.686061e-01	0.238	0.2362	1	2.016182e-03	0.1839	0.1375
E. coli 55989	1	1.035273e-03	0.2324	0.1653	0	1.686061e-01	0.231	0.2432	1	2.016182e-03	0.1858	0.1375
E. coli ABU 83972	1	2.109714e-03	0.2168	0.1676	0	1.686061e-01	0.2377	0.2339	1	2.253333e-03	0.1792	0.1371
E. coli B str. REL606	1	2.049412e-03	0.2148	0.1653	0	1.942353e-01	0.2186	0.2432	1	4.768780e-03	0.1682	0.1375
E. coli BL21-Gold(DE3)pLysS AG	1	2.277600e-02	0.1742	0.1617	0	1.686061e-01	0.2275	0.2432	1	7.599091e-03	0.1599	0.1354
E. coli BL21(DE3) [iECD1391]	1	1.254783e-02	0.1846	0.1624	0	1.686061e-01	0.2214	0.241	1	3.598947e-03	0.1692	0.1354
E. coli DH1 [iEcDH1 1363]	1	2.015000e-03	0.216	0.1653	0	1.686061e-01	0.2451	0.2432	1	2.016182e-03	0.1856	0.1375
E. coli DH1 [iECDH1ME8569 1439]	1	1.028444e-02	0.1892	0.1638	0	3.132683e-01	0.2041	0.2504	1	2.677447e-02	0.1474	0.138
E. coli E24377A	1	1.035273e-03	0.2298	0.1661	0	1.686061e-01	0.2373	0.241	1	2.016182e-03	0.1855	0.1375
E. coli ED1a	1	2.777561e-03	0.2126	0.1684	0	2.051111e-01	0.2036	0.2295	1	6.277143e-03	0.164	0.1367
E. coli O157:H7	1	2.015000e-03	0.219	0.1668	0	1.686061e-01	0.2409	0.2385	1	2.166667e-03	0.1816	0.1375
E. coli HS	1	1.516320e-03	0.2214	0.1653	0	1.686061e-01	0.2237	0.2432	1	2.661176e-03	0.1759	0.1375
E. coli NA114	1	1.102400e-04	0.2734	0.17	0	1.686061e-01	0.2188	0.2317	1	3.786667e-03	0.1716	0.138
E. coli O103:H2 str. 12009	1	1.035273e-03	0.2298	0.1661	0	2.463158e-01	0.2065	0.241	1	2.298065e-03	0.1782	0.1375
E. coli O111:H- str. 11128	1	1.275130e-03	0.2243	0.1653	0	1.686061e-01	0.2394	0.2432	1	2.016182e-03	0.1828	0.1375
E. coli O26:H11 str. 11368	1	1.035273e-03	0.2292	0.1653	0	1.686061e-01	0.2252	0.2432	1	2.016182e-03	0.1833	0.1375
E. coli IHE3034	1	2.777561e-03	0.2128	0.1684	0	1.686061e-01	0.2206	0.2317	1	3.246486e-03	0.1726	0.1371
E. coli ATCC 8739	1	1.516320e-03	0.2228	0.1661	0	1.686061e-01	0.2264	0.241	1	2.253333e-03	0.18	0.1375
E. coli 536	1	2.600000e-03	0.2143	0.1684	0	1.686061e-01	0.2346	0.2317	1	2.270667e-03	0.1782	0.1371
E. coli O157:H7 str. Sakai	1	2.015000e-03	0.2189	0.1668	0	1.686061e-01	0.232	0.2385	1	2.166667e-03	0.1812	0.1375
E. coli S88	1	2.049412e-03	0.2175	0.1676	0	1.686061e-01	0.2164	0.2339	1	3.075429e-03	0.1737	0.1371
E. coli SE11	1	1.035273e-03	0.2316	0.1653	0	1.686061e-01	0.2247	0.2432	1	2.016182e-03	0.1852	0.1375
E. coli SE15	1	3.113500e-04	0.2168	0.1435	0	6.770612e-01	0.1385	0.213	1	2.453750e-03	0.1543	0.1197
E. coli SMS-3-5	1	8.377778e-04	0.2376	0.1646	0	1.686061e-01	0.2341	0.2455	1	2.016182e-03	0.1835	0.1375
E. coli O157:H7 str. TW14359	1	2.557838e-03	0.2138	0.1676	0	1.686061e-01	0.235	0.2362	1	2.253333e-03	0.1799	0.1375
E. coli UMN026	1	1.035273e-03	0.2277	0.1653	0	2.520000e-01	0.2047	0.241	1	2.270667e-03	0.1785	0.1371
E. coli W [iECW_1372]	1	1.035273e-03	0.2328	0.1646	0	1.686061e-01	0.2412	0.2479	1	2.016182e-03	0.196	0.138
E. coli KO11FL	1	1.035273e-03	0.2296	0.1661	0	1.686061e-01	0.2408	0.2432	1	2.016182e-03	0.1876	0.138
E. coli ETEC H10407	1	1.035273e-03	0.2303	0.1653	0	1.686061e-01	0.2247	0.2432	1	2.016182e-03	0.1843	0.1375

E. coli O55:H7 str. CB9615	1	1.035273e-03	0.2293	0.1668	0	1.686061e-01	0.2409	0.2385	1	2.016182e-03	0.192	0.1375
E. coli K-12 MG1655 [iJO1366]	1	2.626000e-04	0.2198	0.143	0	6.770612e-01	0.1404	0.2164	1	2.016182e-03	0.1596	0.12
E. coli K-12 MG1655 [iJR904]	1	4.892727e-03	0.237	0.1945	0	9.510000e-01	0.113	0.2295	1	3.412500e-02	0.1565	0.1498
E. coli LF82	1	1.929200e-07	0.332	0.1676	0	1.686061e-01	0.2357	0.2339	1	1.071200e-07	0.275	0.1371
Mycobacterium tuberculosis H37Rv	1	1.899592e-02	0.1772	0.1617	0	9.043922e-01	0.1007	0.1865	0	2.950000e-01	0.097	0.123
E. coli O83:H1 str. NRG 857C	1	2.325556e-03	0.2143	0.1668	0	1.686061e-01	0.2375	0.2362	1	2.505455e-03	0.1763	0.1371
Shigella flexneri 2a str. 2457T	1	4.462326e-03	0.2014	0.1646	0	2.463158e-01	0.1987	0.2317	1	1.277391e-02	0.1533	0.135
Shigella flexneri 5 str. 8401	1	3.912381e-03	0.2042	0.1653	0	1.686061e-01	0.2328	0.2317	1	1.128978e-02	0.1554	0.1354
E. coli UM146	1	2.015000e-03	0.2187	0.1676	0	1.686061e-01	0.269	0.2362	1	2.016182e-03	0.184	0.1375
E. coli UMNK88	1	1.332500e-02	0.1867	0.1653	0	4.739091e-01	0.1799	0.2432	0	1.799200e-01	0.1185	0.1375
E. coli UTI89	1	1.802000e-03	0.2221	0.1676	0	1.686061e-01	0.2241	0.2339	1	2.270667e-03	0.1782	0.1371
Klebsiella pneumoniae MGH78578	1	1.332500e-02	0.2798	0.2479	0	7.581600e-01	0.1751	0.2847	1	4.499592e-02	0.1922	0.1887
Bacillus subtilis str. 168	1	1.035273e-03	0.2624	0.1898	0	6.367111e-01	0.1895	0.2775	1	2.016182e-03	0.218	0.1583
E. coli O157:H7 str. EDL933	1	1.035273e-03	0.2302	0.1668	0	1.686061e-01	0.2418	0.2385	1	2.016182e-03	0.1923	0.1375
Salmonella Typhimurium str. LT2	1	2.464800e-04	0.2988	0.1922	0	1.686061e-01	0.2112	0.2218	1	2.016182e-03	0.2048	0.1465

Table S15: Operons play an important role in the clustering of strictly essential genes. In this analysis we divided the set of strictly non-essential genes into two groups: *i)* those belonging to an operon, and *ii)* those not belonging to an operon. Using Kuiper's test, we examined the clustering of *i)* the first group of strictly essential genes alone (columns 2-5; labeled as red in the first row), ii) the second group of strictly essential genes alone (columns 6-9; labeled as blue in the first row), and *iii)* all strictly essential genes together (columns 10-13; labeled as red in the first row). Each row corresponds to a bacterial species or strain. The first column is the species (strain) name, and in each of the three set of four columns, from left to right, columns show whether the null hypothesis of a uniform distribution of strictly essential genes is rejected by Kuiper's test (1) or not (0), the P-value of the test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 49 among the 52 genomes the null hypothesis is rejected, i.e., essential genes are significantly clustered. We consider a metabolic gene strictly essential, if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable. In all the 52 genomes used in this analysis, the strictly essential genes belonging to an operon were significantly clustered but strictly essential genes not belonging to an operon were not significantly clustered.

Species(strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
E.coli K-12 MG1655 [iAF1260]	0	2.624186e-01	0.2023	0.2432
Methanosarcina barkeri str. Fusaro	0	6.658039e-01	0.2341	0.3627
Geobacter metallireducens GS-15	0	5.210833e-01	0.1446	0.204
E. coli APEC O1	0	1.935556e-01	0.2045	0.2295
E. coli BL21(DE3) [iB21 1397]	0	3.341277e-01	0.1717	0.2182
E. coli BW2952	0	1.917500e-01	0.2089	0.204
E. coli CFT073	0	1.917500e-01	0.2011	0.2182
E. coli O127:H6	0	2.253333e-01	0.1984	0.2317
E. coli 042	0	1.917500e-01	0.2117	0.2295
E. coli 55989	0	1.917500e-01	0.2186	0.2275
E. coli ABU 83972	0	1.917500e-01	0.2179	0.2362
E. coli B str. REL606	0	1.970526e-01	0.2016	0.2275
E. coli BL21-Gold(DE3)pLysS AG	0	2.765455e-01	0.1857	0.2255
E. coli BL21(DE3) [iECD1391]	0	3.085333e-01	0.1815	0.2255
E. coli DH1 [iEcDH1 1363]	0	1.917500e-01	0.2088	0.2255
E. coli DH1 [iECDH1ME8569 1439]	0	1.917500e-01	0.2097	0.2255
E. coli E24377A	0	1.917500e-01	0.2116	0.2295
E. coli ED1a	0	1.917500e-01	0.2118	0.2317
E. coli O157:H7	0	1.927059e-01	0.2076	0.2295
E. coli HS	0	2.184000e-01	0.1972	0.2275
E. coli NA114	0	1.917500e-01	0.2189	0.2339
E. coli O103:H2 str. 12009	0	1.917500e-01	0.2167	0.2275
E. coli O111:H- str. 11128	0	1.917500e-01	0.2278	0.2275
E. coli O26:H11 str. 11368	0	1.917500e-01	0.224	0.2275
E. coli IHE3034	0	1.917500e-01	0.2146	0.2362
E. coli ATCC 8739	0	1.917500e-01	0.2111	0.2295
E. coli 536	0	1.917500e-01	0.2125	0.2339
E. coli O157:H7 str. Sakai	0	1.917500e-01	0.2141	0.2295
E. coli S88	0	1.935556e-01	0.2113	0.2362
E. coli SE11	0	1.917500e-01	0.2097	0.2275
E. coli SE15	0	1.917500e-01	0.1936	0.2068
E. coli SMS-3-5	0	2.184000e-01	0.1971	0.2275
E. coli O157:H7 str. TW14359	0	1.917500e-01	0.2142	0.2275
E. coli UMN026	0	1.917500e-01	0.2178	0.2295
E. coli W [iECW_1372]	0	1.917500e-01	0.219	0.2255
E. coli KO11FL	0	1.917500e-01	0.2267	0.2295
E. coli ETEC H10407	0	1.917500e-01	0.2186	0.2275
E. coli O55:H7 str. CB9615	0	1.917500e-01	0.2168	0.2295
E. coli K-12 MG1655 [iJO1366]	0	1.917500e-01	0.202	0.2054
E. coli K-12 MG1655 [iJR904]	0	3.097391e-01	0.2287	0.2847
E. coli LF82	0	1.917500e-01	0.2189	0.2339
Mycobacterium tuberculosis H37Rv	0	9.760000e-01	0.1062	0.2295
E. coli O83:H1 str. NRG 857C	0	1.970526e-01	0.2051	0.2317
Shigella flexneri 2a str. 2457T	0	1.917500e-01	0.2018	0.2218
Shigella flexneri 5 str. 8401	0	1.917500e-01	0.2103	0.2295
E. coli UM146	0	1.927059e-01	0.2127	0.2362
E. coli UMNK88	0	1.917500e-01	0.2213	0.2275
E. coli UTI89	0	2.244878e-01	0.1991	0.2317
Klebsiella pneumoniae_MGH78578	0	6.604000e-01	0.2139	0.327
Bacillus subtilis str. 168	0	6.604000e-01	0.1885	0.2886
E. coli O157:H7 str. EDL933	0	1.917500e-01	0.2133	0.2295
Salmonella Typhimurium str. LT2	0	1.917500e-01	0.2028	0.2228

Table S16: Clustering of the strictly essential operons. Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, whether the null hypothesis of a uniform distribution of strictly essential operons is rejected by Kuiper's test (1) or not (0), the P-value of Kuiper's test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In none of the genomes the null hypothesis is rejected, i.e., essential operons are not significantly clustered in any genome. We consider an operon strictly essential if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species(strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
E.coli K-12 MG1655 [iAF1260]	0	2.032727e-01	0.1597	0.1752
Methanosarcina barkeri str. Fusaro	0	6.791200e-01	0.2341	0.3627
Geobacter metallireducens GS-15	0	5.104490e-01	0.1446	0.204
E. coli APEC O1	0	2.094444e-01	0.1551	0.172
E. coli BL21(DE3) [iB21 1397]	0	2.032727e-01	0.1598	0.1638
E. coli BW2952	0	2.032727e-01	0.154	0.161
E. coli CFT073	0	2.032727e-01	0.2011	0.2182
E. coli O127:H6	0	2.600000e-01	0.1451	0.1736
E. coli 042	0	2.032727e-01	0.1579	0.1708
E. coli 55989	0	2.032727e-01	0.173	0.1708
E. coli ABU 83972	0	2.032727e-01	0.1541	0.172
E. coli B str. REL606	0	2.148421e-01	0.1501	0.172
E. coli BL21-Gold(DE3)pLysS AG	0	2.032727e-01	0.1558	0.1728
E. coli BL21(DE3) [iECD1391]	0	2.032727e-01	0.1604	0.172
E. coli DH1 [iEcDH1 1363]	0	2.032727e-01	0.162	0.17
E. coli DH1 [iECDH1ME8569 1439]	0	2.032727e-01	0.1577	0.17
E. coli E24377A	0	2.032727e-01	0.1697	0.1708
E. coli ED1a	0	2.784889e-01	0.1384	0.1692
E. coli O157:H7	0	2.032727e-01	0.1754	0.1736
E. coli HS	0	2.741818e-01	0.1406	0.1708
E. coli NA114	0	2.032727e-01	0.1668	0.1728
E. coli O103:H2 str. 12009	0	2.032727e-01	0.1566	0.172
E. coli O111:H- str. 11128	0	2.032727e-01	0.1641	0.1736
E. coli O26:H11 str. 11368	0	2.032727e-01	0.1717	0.1708
E. coli IHE3034	0	2.032727e-01	0.1548	0.172
E. coli ATCC 8739	0	2.094444e-01	0.1529	0.1728
E. coli 536	0	2.148421e-01	0.1499	0.1708
E. coli O157:H7 str. Sakai	0	2.032727e-01	0.1657	0.1728
E. coli S88	0	2.210000e-01	0.1483	0.172
E. coli SE11	0	2.032727e-01	0.1627	0.1708
E. coli SE15	0	2.032727e-01	0.1494	0.1583
E. coli SMS-3-5	0	2.032727e-01	0.1565	0.1684
E. coli O157:H7 str. TW14359	0	2.032727e-01	0.1695	0.1728
E. coli UMN026	0	2.032727e-01	0.1596	0.1708
E. coli W [iECW_1372]	0	2.032727e-01	0.1539	0.1708
E. coli KO11FL	0	2.032727e-01	0.1669	0.1728
E. coli ETEC H10407	0	2.032727e-01	0.1613	0.1708
E. coli O55:H7 str. CB9615	0	2.200000e-01	0.1483	0.1708
E. coli K-12 MG1655 [iJO1366]	0	2.032727e-01	0.1517	0.1596
E. coli K-12 MG1655 [iJR904]	0	3.783830e-01	0.145	0.1887
E. coli LF82	0	2.511220e-01	0.1557	0.1827
Mycobacterium tuberculosis H37Rv	0	9.900000e-01	0.1062	0.2295
E. coli O83:H1 str. NRG 857C	0	2.032727e-01	0.1644	0.172
Shigella flexneri 2a str. 2457T	0	3.526957e-01	0.1396	0.1786
Shigella flexneri 5 str. 8401	0	2.648372e-01	0.1527	0.1834
E. coli UM146	0	2.032727e-01	0.1612	0.172
E. coli UMNK88	0	2.032727e-01	0.1861	0.1728
E. coli UTI89	0	2.094444e-01	0.1508	0.17
Klebsiella pneumoniae MGH78578	0	4.669167e-01	0.1467	0.2011
Bacillus subtilis str. 168	0	9.900000e-01	0.0833	0.1922
E. coli O157:H7 str. EDL933	0	2.032727e-01	0.1711	0.1728
Salmonella Typhimurium str. LT2	0	2.032727e-01	0.1614	0.1703

Table S17: Clustering of the conditionally essential operons. Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, whether the null hypothesis of uniform distribution of conditionally essential genes is rejected by Kuiper's test (1) or not (0), the P-value of the test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In none of the 55 genomes the null hypothesis is rejected, i.e., conditionally essential operons are not significantly clustered in any genome. We consider an operon conditionally essential, if its deletion abolishes viability on at least one carbon source.

	Among functional subsystem pairs		Among synthetic lethal gene pairs	
Strain(species)	Median of D_{min}	Fraction of short-range pairs	Maximum D_{min}	Fraction of gene pairs in the same subsystem
E.coli K-12 MG1655 [iAF1260]	3	0.9	5	0.4
Methanosarcina barkeri str. Fusaro	3	0.92	7	0.23
Geobacter metallireducens GS-15	4	0.88	1	0.39
E. coli APEC O1	10	0.74	5	0.63
E. coli BL21(DE3) [iB21 1397]	10	0.74	5	0.4
E. coli BW2952	8	0.78	5	0.6
E. coli CFT073	10	0.76	10	0.42
E. coli O127:H6	13	0.71	11	0.63
E. coli 042	9	0.76	5	0.62
E. coli 55989	9	0.78	5	0.62
E. coli ABU 83972	10	0.75	5	0.63
E. coli B str. REL606	10	0.74	5	0.62
E. coli BL21-Gold(DE3)pLysS AG	10	0.73	5	0.4
E. coli BL21(DE3) [iECD1391]	10	0.75	5	0.4
E. coli DH1 [iEcDH1 1363]	8	0.78	5	0.62
E. coli DH1 [iECDH1ME8569 1439]	9	0.75	5	0.61
E. coli E24377A	11	0.74	11	0.63
E. coli ED1a	9	0.77	5	0.63
E. coli O157:H7	8	0.79	5	0.62
E. coli HS	12	0.73	5	0.62
E. coli NA114	8	0.78	6	0.63
E. coli O103:H2 str. 12009	11	0.75	5	0.61
E. coli O111:H- str. 11128	10	0.74	4	0.62
E. coli O26:H11 str. 11368	11	0.75	5	0.62
E. coli IHE3034	9	0.76	4	0.63
E. coli ATCC 8739	11	0.73	6	0.62
E. coli 536	8	0.78	5	0.63
E. coli O157:H7 str. Sakai	8	0.79	5	0.62
E. coli S88	10	0.75	5	0.63
E. coli SE11	11	0.75	5	0.62
E. coli SE15	8	0.78	5	0.6
E. coli SMS-3-5	9	0.76	5	0.62
E. coli O157:H7 str. TW14359	8	0.78	5	0.62
E. coli UMN026	14	0.69	9	0.61
E. coli W [iECW_1372]	11	0.77	6	0.62
E. coli KO11FL	11	0.74	5	0.61
E. coli ETEC H10407	10	0.74	5	0.62
E. coli O55:H7 str. CB9615	9	0.76	5	0.62
E. coli K-12 MG1655 [iJO1366]	2	0.93	5	0.57
E. coli K-12 MG1655 [iJR904]	5	0.84	1	0.3
E. coli LF82	18	0.66	7	0.63
Mycobacterium tuberculosis H37Rv	2	0.98	4	0.12
E. coli O83:H1 str. NRG 857C	14	0.71	9	0.63
Shigella flexneri 2a str. 2457T	10	0.8	1	0.65
Staphylococcus aureus N315	27	0.62	1	0.19
Shigella flexneri 5 str. 8401	9	0.78	1	0.67
E. coli UM146	10	0.75	5	0.62
E. coli UMNK88	12	0.73	7	0.62
E. coli UTI89	11	0.75	4	0.63
E. coli W [iWFL 1372]	11	0.77	6	0.62
E. coli str. K-12 W3110	8	0.78	5	0.62
Klebsiella pneumoniae MGH78578	16	0.7	5	0.46
Bacillus subtilis str. 168	0	1	1	0.5
E. coli O157:H7 str. EDL933	8	0.79	5	0.62
Salmonella Typhimurium str. LT2	9	0.77	1	0.37

Table S18: Repulsion of synthetic lethal gene pairs is not due to the repulsion of subsystem pairs. Each row corresponds to one of the 55 bacterial genomes. Columns, from left to right, show, first, the species (strain) name, second, the median of the $D_{min}(sub_i, sub_j)$, which is defined as the minimum distance (i.e. the number of intervening metabolic genes) between a given pair of subsystems, among all pairs of subsystems, third, the fraction of subsystem pairs (i, j) with $D_{min}(sub_i, sub_j) < 50$, fourth, the maximum $D_{min}(sub_{gene(i)}, sub_{gene(j)})$ among all strictly lethal gene pairs, (i.e. $gene(i)$, $gene(j)$), where $sub_{gene(i)}$ indicates the subsystem to which $gene(i)$ belongs, and, fifth, the fraction of synthetic lethal gene pairs in which both genes belong to the same subsystem.

	Among metabolic pathway pairs		Among synthetic lethal gene pairs	
Strain(species)	Median of D_{min}	Fraction of short-range pairs	Maximum D_{min}	Fraction of gene pairs in the same metabolic pathway
E.coli K-12 MG1655 [iAF1260]	4	0.88	1	0.89
Methanosarcina barkeri str. Fusaro	2	0.93	1	0.69
Geobacter metallireducens GS-15	3	0.89	1	0.79
E. coli APEC O1	4	0.89	1	0.63
E. coli BL21(DE3) [iB21 1397]	4	0.88	1	0.75
E. coli BW2952	4	0.88	1	0.65
E. coli O127:H6	4	0.88	1	0.63
E. coli 55989	5	0.87	1	0.63
E. coli ABU 83972	4	0.88	1	0.63
E. coli B str. REL606	4	0.88	1	0.63
E. coli BL21-Gold(DE3)pLysS AG	4	0.88	1	0.75
E. coli BL21(DE3) [iECD1391]	4	0.88	1	0.75
E. coli NA114	5	0.87	1	0.63
E. coli O103:H2 str. 12009	4	0.87	1	0.61
E. coli O111:H- str. 11128	5	0.87	1	0.63
E. coli O26:H11 str. 11368	5	0.87	1	0.63
E. coli IHE3034	4	0.88	1	0.63
E. coli ATCC 8739	4	0.88	1	0.63
E. coli O157:H7 str. Sakai	4	0.87	1	0.63
E. coli SMS-3-5	5	0.86	1	0.63
E. coli O157:H7 str. TW14359	4	0.87	1	0.63
E. coli W [iECW_1372]	5	0.86	1	0.64
E. coli KO11FL	5	0.87	1	0.64
E. coli O55:H7 str. CB9615	4	0.87	1	0.63
E. coli K-12 MG1655 [iJO1366]	4	0.88	1	0.63
E. coli K-12 MG1655 [iJR904]	4	0.87	1	0.9
E. coli LF82	8	0.79	1	0.63
Mycobacterium tuberculosis H37Rv	2	0.92	1	0.59
E. coli O83:H1 str. NRG 857C	4	0.88	1	0.63
Shigella flexneri 2a str. 2457T	3	0.9	1	0.65
Shigella flexneri 5 str. 8401	3	0.89	1	0.58
E. coli UM146	4	0.89	1	0.63
E. coli UMNK88	4	0.87	1	0.63
E. coli W [iWFL 1372]	5	0.86	1	0.64
Klebsiella pneumoniae MGH78578	4	0.88	1	0.95
Bacillus subtilis str. 168	4	0.89	1	0.62
E. coli O157:H7 str. EDL933	4	0.86	1	0.58
Salmonella Typhimurium str. LT2	4	0.86	1	0.8

Table S19: Repulsion of synthetic lethal gene pairs is not due to the repulsion of metabolic pathway pairs. Each row corresponds to one of 38 bacterial genomes among the 55 genomes used in this study, whose pathway information are available in KEGG database. Columns, from left to right, show, first, the species (strain) name, second, the median of the $D_{min}(P_i, P_j)$, which is defined as the minimum distance (i.e. the number of intervening metabolic genes) between a given pair of metabolic pathways, among all pairs of pathways, third, the fraction of metabolic pathway pairs (i, j) with $D_{min}(P_i, P_j) < 50$, fourth, the maximum $D_{min}(P_{gene(i)}, P_{gene(j)})$ among all strictly lethal gene pairs, (i.e. $gene(i)$, $gene(j)$), where $P_{gene(i)}$ indicates the metabolic pathway to which $gene(i)$ belongs, and, fifth, the fraction of synthetic lethal gene pairs in which both genes belong to the same metabolic pathway.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
Methanosarcina barkeri str. Fusaro	273	160	169	90	0.9940	0.9087	0
Geobacter metallireducens GS-15	271	453	83	180	0.9030	1.2974	0
E. coli BL21(DE3) [iB21 1397]	365	673	89	210	0.9030	1.2797	0
E. coli BW2952	434	594	119	181	0.9280	1.1113	0
E. coli CFT073	463	580	115	149	1.0000	1.0343	0
E. coli O127:H6	438	565	111	170	0.9170	1.1873	0
E. coli 042	410	587	130	187	1.0000	1.0047	0
E. coli 55989	428	599	118	185	0.9280	1.1202	0
E. coli ABU 83972	428	578	133	181	1.0000	1.0077	0
E. coli B str. REL606	439	588	121	181	0.9280	1.1168	0
E. coli BL21-Gold(DE3)pLysS AG	452	599	122	181	0.9280	1.1195	0
E. coli BL21(DE3) [iECD1391]	470	612	98	153	0.9170	1.199	0
E. coli DH1 [iEcDH1 1363]	443	616	117	187	0.9280	1.1494	0
E. coli O157:H7	436	534	117	175	0.9030	1.2212	0
E. coli HS	428	584	119	190	0.9170	1.1701	0
E. coli NA114	428	568	130	175	1.0000	1.0144	0
E. coli O103:H2 str. 12009	444	573	118	192	0.9030	1.2608	0
E. coli O111:H- str. 11128	442	577	120	189	0.9030	1.2065	0
E. coli O26:H11 str. 11368	438	614	123	180	1.0000	1.0439	0
E. coli IHE3034	448	565	127	164	1.0000	1.0239	0
E. coli ATCC 8739	440	612	127	189	1.0000	1.0699	0
E. coli 536	436	559	136	178	1.0000	1.0208	0
E. coli O157:H7 str. Sakai	445	568	115	173	0.9170	1.1786	0
E. coli S88	440	580	119	166	1.0000	1.0582	0
E. coli SE11	427	609	124	188	1.0000	1.063	0
E. coli SE15	431	597	125	174	1.0000	1.0049	0
E. coli SMS-3-5	419	603	126	199	0.9870	1.0974	0
E. coli O157:H7 str. TW14359	445	551	120	183	0.9030	1.2316	0
E. coli W [iECW_1372]	458	603	128	183	0.9940	1.0859	0
E. coli KO11FL	438	609	120	187	0.9280	1.1208	0
E. coli ETEC H10407	435	581	125	192	0.9280	1.15	0
E. coli O55:H7 str. CB9615	407	561	129	186	1.0000	1.0461	0
E. coli LF82	439	575	121	167	1.0000	1.0537	0
E. coli O83:H1 str. NRG 857C	439	582	120	170	1.0000	1.0686	0
Shigella flexneri 2a str. 2457T	416	472	133	167	0.9280	1.1067	0
Shigella flexneri 5 str. 8401	406	454	138	186	0.9030	1.2053	0
E. coli UM146	447	588	124	160	1.0000	0.9809	0
E. coli UMNK88	432	583	135	203	0.9280	1.1142	0
E. coli UTI89	424	589	123	174	1.0000	1.0183	0
Klebsiella pneumoniae MGH78578	393	511	156	169	0.9030	0.8332	0
Bacillus subtilis str. 168	250	403	72	119	1.0000	1.0253	0
Salmonella Typhimurium str. LT2	418	554	124	175	1.0000	1.0648	0

Table S20: Horizontal gene transfer and operons. Each row corresponds to a given species or strain. Columns, from left to right, show species (strain) names, the number of metabolic genes that are neither HGT-acquired nor belong to an operon $(--)$, the number of metabolic genes that are not HGT-acquired but do belong to an operon $(-+)$, the number of metabolic genes that are HGT-acquired, but do not belong to an operon $(+-)$, the number of metabolic genes that are both HGT-acquired and belong to an operon $(++)$, the P value of a Fisher's exact test carried out for this data, the odds ratio (defined as the odds of being HGT-acquired for operonic metabolic genes divided by the odds of being HGT-acquired for non-operonic metabolic genes), and whether the null hypothesis of a lack of association between horizontal gene transfer and being on an operon is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). The data shows that no significant association between horizontally transferred genes and genes belonging to operons in any of our study genomes.

Species(strains)	number of operons	average number of genes per opeon	average robustness to operon deletion (strict definition)	average (normalized) robustness of genome (strict definition)	excess robustness to operon deletion (strict definition)	R(tandem)/R(random) (strict definition)	average robustness to operon deletion (moderate definition)	average (normalized) robustness of genome (moderate definition)	excess robustness to operon deletion (moderate definition)	R(tandem)/R(random) (moderate definition)
E.coli K-12 MG1655 [iAF1260]	241	3.07	0.6	0.56	1.08	1.24	0.8	0.77	1.03	1.12
Methanosarcina barkeri str. Fusaro	91	2.75	0.77	0.62	1.23	1.09	0.77	0.62	1.23	1.09
Geobacter metallireducens GS-15	203	3.12	0.66	0.56	1.17	1.3	0.66	0.56	1.17	1.3
E. coli APEC O1	247	2.94	0.59	0.57	1.04	1.19	0.79	0.76	1.04	1.08
E. coli BL21(DE3) [iB21 1397]	277	3.19	0.6	0.57	1.07	1.23	0.78	0.75	1.03	1.11
E. coli BW2952	247	3.14	0.54	0.52	1.03	1.26	0.72	0.69	1.03	1.13
E. coli CFT073	260	2.8	0.77	0.72	1.06	1.13	0.77	0.72	1.06	1.13
E. coli O127:H6	244	3.01	0.6	0.57	1.05	1.24	0.78	0.76	1.03	1.12
E. coli 042	252	3.07	0.6	0.56	1.06	1.23	0.78	0.76	1.03	1.12
E. coli 55989	257	3.05	0.61	0.57	1.06	1.19	0.78	0.76	1.02	1.09
E. coli ABU 83972	247	3.07	0.6	0.57	1.04	1.17	0.79	0.76	1.04	1.1
E. coli B str. REL606	256	3	0.61	0.57	1.06	1.23	0.78	0.77	1.02	1.12
E. coli BL21-Gold(DE3)pLysS AG	256	3.05	0.61	0.58	1.06	1.21	0.78	0.76	1.02	1.12
E. coli BL21(DE3) [iECD1391]	253	3.02	0.6	0.57	1.06	1.22	0.77	0.76	1.02	1.12
E. coli DH1 [iEcDH1 1363]	260	3.09	0.61	0.58	1.04	1.2	0.78	0.77	1.01	1.1
E. coli DH1 [iECDH1ME8569 1439]	255	3.1	0.6	0.61	0.98	1.2	0.78	0.79	0.99	1.11
E. coli E24377A	260	2.98	0.61	0.58	1.05	1.18	0.79	0.77	1.02	1.1
E. coli ED1a	242	3	0.57	0.56	1.02	1.23	0.78	0.76	1.03	1.13
E. coli O157:H7	240	2.95	0.59	0.56	1.05	1.22	0.77	0.76	1.02	1.11
E. coli HS	253	3.06	0.6	0.57	1.06	1.24	0.78	0.76	1.02	1.11
E. coli NA114	245	3.03	0.6	0.57	1.05	1.24	0.78	0.76	1.03	1.1
E. coli O103:H2 str. 12009	250	3.06	0.6	0.58	1.04	1.23	0.78	0.76	1.02	1.11
E. coli O111:H- str. 11128	251	3.05	0.61	0.58	1.05	1.24	0.78	0.76	1.02	1.1
E. coli O26:H11 str. 11368	261	3.04	0.61	0.58	1.06	1.21	0.79	0.77	1.02	1.09
E. coli IHE3034	242	3.01	0.59	0.57	1.04	1.22	0.79	0.76	1.03	1.12
E. coli ATCC 8739	258	3.1	0.62	0.58	1.06	1.25	0.79	0.77	1.02	1.12
E. coli 536	248	2.97	0.59	0.57	1.04	1.18	0.79	0.76	1.03	1.1
E. coli O157:H7 str. Sakai	248	2.99	0.6	0.57	1.05	1.2	0.78	0.76	1.02	1.1
E. coli S88	248	3.01	0.6	0.57	1.04	1.22	0.79	0.76	1.04	1.11
E. coli SE11	259	3.08	0.61	0.58	1.05	1.23	0.78	0.77	1.02	1.1
E. coli SE15	251	3.07	0.53	0.51	1.04	1.24	0.73	0.7	1.05	1.13
E. coli SMS-3-5	260	3.08	0.6	0.57	1.05	1.22	0.78	0.76	1.03	1.09
E. coli O157:H7 str. TW14359	245	3	0.6	0.57	1.04	1.21	0.77	0.76	1.01	1.11
E. coli UMN026	259	3.05	0.61	0.57	1.08	1.22	0.79	0.76	1.03	1.12
E. coli W [iECW_1372]	259	3.03	0.61	0.59	1.04	1.18	0.78	0.77	1.01	1.11
E. coli KO11FL	258	3.09	0.62	0.58	1.06	1.21	0.79	0.77	1.03	1.11
E. coli ETEC H10407	257	3.01	0.61	0.58	1.04	1.2	0.78	0.77	1.02	1.09
E. coli O55:H7 str. CB9615	247	3.02	0.59	0.55	1.07	1.2	0.78	0.76	1.03	1.11
E. coli K-12 MG1655 [iJO1366]	262	3.12	0.56	0.52	1.07	1.26	0.74	0.71	1.04	1.13
E. coli K-12 MG1655 [iJR904]	181	2.87	0.55	0.43	1.28	1.29	0.81	0.72	1.11	1.12
E. coli LF82	246	3.02	0.6	0.58	1.03	1.22	0.78	0.77	1.01	1.13
Mycobacterium tuberculosis H37Rv	122	2.82	0.55	0.48	1.15	1.2	0.55	0.48	1.15	1.2

E. coli O83:H1 str. NRG 857C	248	3.03	0.6	0.58	1.03	1.21	0.78	0.76	1.03	1.1
Shigella flexneri 2a str. 2457T	218	2.93	0.58	0.55	1.04	1.22	0.73	0.73	1	1.12
Shigella flexneri 5 str. 8401	218	2.94	0.6	0.53	1.13	1.2	0.75	0.81	0.93	1.08
E. coli UM146	249	3	0.6	0.55	1.08	1.18	0.79	0.73	1.08	1.12
E. coli UMNK88	259	3.03	0.62	0.57	1.07	1.19	0.78	0.76	1.03	1.1
E. coli UTI89	259	2.95	0.61	0.57	1.06	1.16	0.79	0.75	1.06	1.03
Klebsiella pneumoniae MGH78578	227	3	0.68	0.57	1.19	1.17	0.89	0.76	1.16	1.1
Bacillus subtilis str. 168	161	3.24	0.51	0.58	0.88	1.23	0.79	0.76	1.03	1.11
E. coli O157:H7 str. EDL933	253	3.02	0.61	0.58	1.04	1.22	0.78	0.77	1.01	1.11
Salmonella Typhimurium str. LT2	247	2.95	0.66	0.63	1.05	1.16	0.82	0.85	0.96	1.04

Table S21: Operons and deletional robustness. Each row corresponds to a given species or strain. Columns, from left to right, show species or strain names, the total number of operons in the genome, the average number of genes per operon, the average robustness to operon deletion, average normalized robustness to tandem gene deletion (see text S11), excess robustness to operon deletion (i.e. the robustness to operon deletion divided by the average robustness to tandem deletion), excess robustness to tandem deletion ($\bar{R}_{tandem}/\bar{R}_{random}$). In columns 4-7, robustness requires retaining viability on *all* carbon sources. Columns 8 to 11 shows the same information as columns 4-7 respectively but for a less stringent definition of robustness that requires retaining viability on at least one carbon source

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
Methanosarcina barkeri str. Fusaro	357	196	76	63	4.214000e-02	1.5099	1
Geobacter metallireducens GS-15	535	184	189	79	2.303571e-01	1.2153	0
E. coli BL21(DE3) [iB21 1397]	932	241	106	58	7.135313e-05	2.116	1
E. coli BW2952	888	230	140	70	1.190861e-04	1.9304	1
E. coli CFT073	896	207	147	57	4.674872e-03	1.6784	1
E. coli O127:H6	898	226	105	55	1.141343e-04	2.0813	1
E. coli 042	903	252	94	65	4.690909e-06	2.4778	1
E. coli 55989	928	243	99	60	9.910476e-06	2.3145	1
E. coli ABU 83972	910	250	96	64	4.690909e-06	2.4267	1
E. coli B str. REL606	928	242	99	60	9.825500e-06	2.3241	1
E. coli BL21-Gold(DE3)pLysS AG	946	244	105	59	3.583333e-05	2.1785	1
E. coli BL21(DE3) [iECD1391]	967	202	115	49	2.874211e-04	2.0397	1
E. coli DH1 [iEcDH1 1363]	960	244	99	60	6.745625e-06	2.3845	1
E. coli O157:H7	872	231	98	61	9.825500e-06	2.3497	1
E. coli HS	916	246	96	63	4.690909e-06	2.4436	1
E. coli NA114	897	246	99	59	4.716129e-05	2.1731	1
E. coli O103:H2 str. 12009	921	247	96	63	4.690909e-06	2.447	1
E. coli O111:H- str. 11128	919	250	100	59	4.716129e-05	2.1688	1
E. coli O26:H11 str. 11368	953	243	99	60	6.745625e-06	2.3769	1
E. coli IHE3034	914	230	99	61	4.801667e-06	2.4486	1
E. coli ATCC 8739	956	253	96	63	4.690909e-06	2.4797	1
E. coli 536	898	251	97	63	9.825500e-06	2.3237	1
E. coli O157:H7 str. Sakai	913	229	100	59	1.063273e-05	2.3523	1
E. coli S88	920	225	100	60	4.690909e-06	2.4533	1
E. coli SE11	936	253	100	59	3.583333e-05	2.1828	1
E. coli SE15	894	223	134	76	4.690909e-06	2.2737	1
E. coli SMS-3-5	926	262	96	63	9.825500e-06	2.3194	1
E. coli O157:H7 str. TW14359	899	241	97	62	6.204286e-06	2.3843	1
E. coli W [iECW_1372]	962	252	99	59	2.037826e-05	2.2751	1
E. coli KO11FL	951	245	96	62	4.690909e-06	2.5069	1
E. coli ETEC H10407	917	257	99	60	4.003448e-05	2.1625	1
E. coli O55:H7 str. CB9615	869	255	99	60	9.346176e-05	2.0654	1
E. coli LF82	896	224	118	64	2.239583e-05	2.1695	1
E. coli O83:H1 str. NRG 857C	918	233	103	57	4.003448e-05	2.1803	1
Shigella flexneri 2a str. 2457T	784	237	104	63	1.301622e-04	2.0039	1
Shigella flexneri 5 str. 8401	763	257	97	67	8.443636e-05	2.0507	1
E. coli UM146	935	225	100	59	5.226154e-06	2.4518	1
E. coli UMNK88	919	275	96	63	2.528400e-05	2.1931	1
E. coli UTI89	915	235	98	62	4.690909e-06	2.4633	1
E. coli str. K-12 W3110	1141	58	130	29	1.066400e-06	4.3885	1
Klebsiella pneumoniae MGH78578	844	296	60	29	1.803902e-01	1.3782	0
Bacillus subtilis str. 168	561	163	92	28	8.150000e-01	1.0475	0
Salmonella Typhimurium str. LT2	892	239	80	60	1.580250e-06	2.7992	1

Table S22: Horizontal gene transfer and strictly essential genes. Each row corresponds to one of the 43 species strains for which information about horizontally transferred genes (HGT) is available in the HGTree database. Columns, from left to right, show the species (strain) name, the number of metabolic genes that are neither strictly essential nor HGT-acquired $(--)$, the number of metabolic genes that are not strictly essential, but HGT-acquired $(-+)$, the number of metabolic genes that are strictly essential, but are not HGT-acquired $(+-)$, and the number of metabolic genes that are both strictly essential and HGT-acquired $(++)$, the P value of a Fisher exact test on this data, the odds ratio (defined as the odds of being strictly essential among HGT-acquired genes divided by the odds of being strictly essential among non-HGT acquired essential genes), and whether the null hypothesis of a lack of association between a gene's strict essentiality and horizontal transfer is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction(32). In 40 of the 43 species (93.02%) the null hypothesis is rejected. Note that we consider a metabolic gene strictly essential, if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
Methanosarcina barkeri str. Fusaro	357	196	76	63	4.555676e-02	1.5099	1
Geobacter metallireducens GS-15	535	184	189	79	2.418750e-01	1.2153	0
E. coli BL21(DE3) [iB21 1397]	808	203	230	96	1.705667e-03	1.6613	1
E. coli BW2952	764	197	264	103	6.511429e-03	1.5131	1
E. coli CFT073	896	207	147	57	6.511429e-03	1.6784	1
E. coli O127:H6	775	196	228	85	1.625937e-02	1.4741	1
E. coli 042	770	215	227	102	2.131304e-03	1.6093	1
E. coli 55989	799	205	228	98	1.505000e-03	1.6753	1
E. coli ABU 83972	779	221	227	93	1.654848e-02	1.4441	1
E. coli B str. REL606	799	203	228	99	1.237444e-03	1.709	1
E. coli BL21-Gold(DE3)pLysS AG	823	205	228	98	1.124143e-03	1.7256	1
E. coli BL21(DE3) [iECD1391]	842	164	240	87	1.006200e-03	1.8611	1
E. coli DH1 [iEcDH1 1363]	832	208	227	96	1.476333e-03	1.6916	1
E. coli O157:H7	712	193	258	99	2.225882e-02	1.4156	1
E. coli HS	778	207	234	102	1.705667e-03	1.6383	1
E. coli NA114	768	213	228	92	1.625937e-02	1.4549	1
E. coli O103:H2 str. 12009	796	209	221	101	1.053500e-03	1.7406	1
E. coli O111:H- str. 11128	778	204	241	105	1.476333e-03	1.6616	1
E. coli O26:H11 str. 11368	813	205	239	98	2.104611e-03	1.6262	1
E. coli IHE3034	786	198	227	93	2.131304e-03	1.6264	1
E. coli ATCC 8739	830	215	222	101	1.006200e-03	1.7563	1
E. coli 536	767	222	228	92	2.854722e-02	1.3941	1
E. coli O157:H7 str. Sakai	785	193	228	95	1.476333e-03	1.6947	1
E. coli S88	792	194	228	91	2.232692e-03	1.6294	1
E. coli SE11	810	215	226	97	2.131304e-03	1.617	1
E. coli SE15	763	189	265	110	1.237444e-03	1.6758	1
E. coli SMS-3-5	796	223	226	102	2.131304e-03	1.611	1
E. coli O157:H7 str. TW14359	771	206	225	97	2.131304e-03	1.6135	1
E. coli W [iECW_1372]	832	214	229	97	1.709250e-03	1.6468	1
E. coli KO11FL	825	208	222	99	1.006200e-03	1.7688	1
E. coli ETEC H10407	781	219	235	98	8.837241e-03	1.4872	1
E. coli O55:H7 str. CB9615	730	216	238	99	2.260571e-02	1.4058	1
E. coli LF82	712	187	302	101	1.093105e-01	1.2734	0
E. coli O83:H1 str. NRG 857C	792	202	229	88	9.302333e-03	1.5067	1
Shigella flexneri 2a str. 2457T	645	209	243	91	3.513415e-01	1.1557	0
Shigella flexneri 5 str. 8401	639	225	221	99	1.168718e-01	1.2722	0
E. coli UM146	806	194	229	90	2.232692e-03	1.6328	1
E. coli UMNK88	799	235	216	103	2.104611e-03	1.6213	1
E. coli UTI89	787	203	226	94	2.221667e-03	1.6125	1
E. coli str. K-12 W3110	981	48	290	39	5.848000e-04	2.7485	1
Klebsiella pneumoniae MGH78578	608	219	296	106	1	0.9942	0
Bacillus subtilis str. 168	425	131	228	60	3.962143e-01	0.8538	0
Salmonella Typhimurium str. LT2	773	205	199	94	1.006200e-03	1.7811	1

Table S23: Horizontal gene transfer and conditionally essential genes. Each row corresponds to one of the 43 species (strain) for which information about horizontally transferred genes (HGT) is available in the HGTree database. Columns, from left to right, show the species (strain) name, the number of metabolic genes that are neither conditionally essential nor HGT-acquired ($--$), the number of metabolic genes that are not conditionally essential, but HGT-acquired ($-+$), the number of metabolic genes that are conditionally essential, but are not HGT-acquired ($+-$), and the number of metabolic genes that are both conditionally essential and HGT-acquired ($++$), the P value of a Fisher exact test on this data, the odds ratio (defined as the odds of being conditionally essential among HGT acquired metabolic genes divided by the odds of being conditionally essential among non-HGT acquired metabolic genes), and whether the null hypothesis of a lack of association between a gene's conditional essentiality and horizontal transfer is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 37 of the 43 species (86.05%) the null hypothesis is rejected. Note that we consider a metabolic gene conditionally essential, if its deletion abolishes viability on at least one carbon source.

	HGT-acquired essential genes				Non-HGT acquired essential genes				All essential genes			
Species(strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value	Hypothesis rejected	P-value	Kuiper test statistic	Critical value	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
Methanosarcina barkeri str. Fusaro	0	7.870000e-01	0.1285	0.2182	0	3.378571e-01	0.1556	0.2011	0	2.530000e-01	0.1211	0.1492
Geobacter metallireducens GS-15	0	3.135854e-01	0.1517	0.1922	1	7.396000e-04	0.2029	0.1293	1	3.956000e-04	0.1688	0.1078
E. coli BL21(DE3) [iB21 1397]	1	1.310067e-02	0.2581	0.2236	0	2.698250e-01	0.1368	0.1684	1	4.369730e-03	0.1656	0.1354
E. coli BW2952	1	8.600000e-03	0.2635	0.204	0	7.525000e-02	0.1516	0.1465	1	2.037722e-03	0.1694	0.1197
E. coli CFT073	1	6.589750e-03	0.3108	0.2295	0	7.525000e-02	0.1621	0.1465	1	2.037722e-03	0.1687	0.1242
E. coli O127:H6	1	9.794444e-03	0.2796	0.2295	0	1.641818e-01	0.1524	0.1692	1	2.889063e-03	0.1733	0.1371
E. coli 042	1	1.143207e-02	0.2482	0.2114	0	7.525000e-02	0.1879	0.1786	1	2.037722e-03	0.1839	0.1375
E. coli 55989	1	8.600000e-03	0.2747	0.22	0	7.525000e-02	0.1829	0.1744	1	2.037722e-03	0.1858	0.1375
E. coli ABU 83972	1	1.102950e-02	0.2559	0.213	0	1.137964e-01	0.1698	0.1769	1	2.096250e-03	0.1792	0.1371
E. coli B str. REL606	1	1.143207e-02	0.2601	0.22	0	1.193967e-01	0.1654	0.1744	1	4.369730e-03	0.1682	0.1375
E. coli BL21-Gold(DE3)pLysS AG	1	8.600000e-03	0.2794	0.2218	0	3.430000e-01	0.1296	0.1692	1	7.276053e-03	0.1599	0.1354
E. coli BL21(DE3) [iECD1391]	1	1.143207e-02	0.2835	0.2432	0	1.783235e-01	0.1434	0.1617	1	3.326176e-03	0.1692	0.1354
E. coli DH1 [iEcDH1 1363]	1	8.600000e-03	0.2773	0.22	0	7.525000e-02	0.1854	0.1744	1	2.037722e-03	0.1856	0.1375
E. coli O157:H7	1	1.929487e-02	0.2399	0.2182	0	7.525000e-02	0.1946	0.1752	1	2.047619e-03	0.1816	0.1375
E. coli HS	1	1.609429e-02	0.2415	0.2147	0	7.525000e-02	0.1898	0.1769	1	2.494000e-03	0.1759	0.1375
E. coli NA114	1	1.479706e-02	0.2514	0.2218	0	7.525000e-02	0.1981	0.1744	1	3.489143e-03	0.1716	0.138
E. coli O103:H2 str. 12009	1	1.929487e-02	0.2376	0.2147	0	7.525000e-02	0.1921	0.1769	1	2.181852e-03	0.1782	0.1375
E. coli O111:H- str. 11128	1	8.600000e-03	0.2763	0.2218	0	7.525000e-02	0.1884	0.1736	1	2.037722e-03	0.1828	0.1375
E. coli O26:H11 str. 11368	1	1.479706e-02	0.2496	0.22	0	7.525000e-02	0.1804	0.1744	1	2.037722e-03	0.1833	0.1375
E. coli IHE3034	1	9.164375e-03	0.2687	0.2182	0	1.917568e-01	0.1512	0.1744	1	3.010000e-03	0.1726	0.1371
E. coli ATCC 8739	1	8.600000e-03	0.2702	0.2147	0	1.100800e-01	0.1726	0.1769	1	2.096250e-03	0.18	0.1375
E. coli 536	1	1.929487e-02	0.2363	0.2147	0	8.824348e-02	0.177	0.176	1	2.166538e-03	0.1782	0.1371
E. coli O157:H7 str. Sakai	1	8.600000e-03	0.2927	0.2218	0	1.244226e-01	0.1628	0.1736	1	2.047619e-03	0.1812	0.1375
E. coli S88	1	1.143207e-02	0.2613	0.22	0	1.109731e-01	0.1682	0.1736	1	2.871290e-03	0.1737	0.1371
E. coli SE11	1	1.143207e-02	0.2643	0.2218	0	7.525000e-02	0.1871	0.1736	1	2.037722e-03	0.1852	0.1375
E. coli SE15	1	6.589750e-03	0.2715	0.1958	0	1.793714e-01	0.1322	0.1498	1	2.318929e-03	0.1543	0.1197
E. coli SMS-3-5	1	1.050105e-02	0.2595	0.2147	0	1.249688e-01	0.1654	0.1769	1	2.037722e-03	0.1835	0.1375
E. coli O157:H7 str. TW14359	1	1.328839e-02	0.249	0.2164	0	8.824348e-02	0.1776	0.176	1	2.096250e-03	0.1799	0.1375
E. coli W [iECW_1372]	1	8.600000e-03	0.2876	0.2218	0	7.525000e-02	0.1898	0.1744	1	2.037722e-03	0.196	0.138
E. coli KO11FL	1	1.143207e-02	0.2554	0.2164	0	7.525000e-02	0.2024	0.1769	1	2.037722e-03	0.1876	0.138
E. coli ETEC H10407	1	8.600000e-03	0.273	0.22	0	7.525000e-02	0.1814	0.1744	1	2.037722e-03	0.1843	0.1375
E. coli O55:H7 str. CB9615	1	1.143207e-02	0.2569	0.22	0	7.525000e-02	0.1891	0.1744	1	2.037722e-03	0.192	0.1375
E. coli LF82	1	6.020000e-04	0.3511	0.2218	1	9.660667e-03	0.232	0.1728	1	8.858000e-08	0.275	0.1371
E. coli O83:H1 str. NRG 857C	1	9.794444e-03	0.2746	0.2255	0	1.129148e-01	0.1647	0.1708	1	2.357586e-03	0.1763	0.1371
Shigella flexneri 2a str. 2457T	1	1.143207e-02	0.2524	0.2147	0	1.839444e-01	0.1505	0.172	1	1.214750e-02	0.1533	0.135
Shigella flexneri 5 str. 8401	1	1.929487e-02	0.2295	0.2083	0	2.093421e-01	0.1502	0.176	1	1.077205e-02	0.1554	0.1354
E. coli UM146	1	1.143207e-02	0.2626	0.2218	0	1.100800e-01	0.169	0.1736	1	2.037722e-03	0.184	0.1375
E. coli UMNK88	0	4.474048e-01	0.156	0.2147	0	2.425641e-01	0.1469	0.1769	0	1.771190e-01	0.1185	0.1375
E. coli UTI89	1	8.600000e-03	0.2724	0.2164	0	1.193967e-01	0.1659	0.1752	1	2.166538e-03	0.1782	0.1371
E. coli str. K-12 W3110	1	8.600000e-03	0.3942	0.3115	0	7.525000e-02	0.1602	0.1521	1	2.047619e-03	0.1815	0.1375
Klebsiella pneumoniae_MGH78578	1	1.479706e-02	0.3546	0.3115	0	3.125366e-01	0.1842	0.2339	1	4.446829e-02	0.1922	0.1887
Bacillus subtilis str. 168	0	2.440250e-01	0.2629	0.3163	1	4.278500e-03	0.2574	0.1804	1	2.037722e-03	0.218	0.1583
Salmonella Typhimurium str. LT2	1	6.589750e-03	0.2977	0.22	0	8.804762e-02	0.1967	0.1933	1	2.037722e-03	0.2048	0.1465

Table S24: Horizontal gene transfer plays an active role in the clustering of the strictly essential genes. In this analysis we divided the set of strictly non-essential genes into two groups: *i)* those acquired via horizontal gene transfer, and *ii)* those not acquired through horizontal gene transfer. Using Kuiper's test, we examined the clustering of *i)* the first group of strictly essential genes alone (columns 2-5; labeled red in the first row), ii) the second group of strictly essential genes alone (columns 6-9; labeled blue in the first row), and *iii)* all strictly essential genes together (columns 10-13; labeled black in the first row). Each row corresponds to a bacterial species (strain). The first column is the species or strain name, and in each of the three sets of four columns, from left to right, columns show whether the null hypothesis of uniform distribution of strictly essential genes is rejected by Kuiper's test (1) or not (0), the P-value of the test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). In 38 out of the 43 genomes (88.37%, colored blue) used in this analysis, the strictly essential genes acquired by horizontal gene transfer are significantly clustered (first group), but the strictly essential genes not acquired by horizontal gene transfer (second group) are not significantly clustered. Only in two genomes (4.65%, colored red) are the strictly essential genes of the second group significantly clustered but genes in the first group are not. In only one genome neither of the groups of strictly essential genes are significantly clustered (2.32%, colored green). Finally, in 2 out of the 43 genomes (4.65%, colored black) the strictly essential genes of both groups are significantly clustered. Note that we consider a metabolic gene strictly essential, if its deletion abolishes viability on all carbon sources on which the wild-type metabolism is viable.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
Methanosarcina barkeri str. Fusaro	18	43	21	51	1	1.0166	0
Geobacter metallireducens GS-15	57	22	117	61	0.3858	1.3508	0
E. coli BL21(DE3) [iB21 1397]	44	14	79	25	1	0.9946	0
E. coli BW2952	47	23	98	40	0.6325	0.8341	0
E. coli CFT073	35	20	78	60	0.4198	1.3462	0
E. coli O127:H6	35	20	71	32	0.5943	0.7887	0
E. coli 042	40	25	65	27	0.3017	0.6646	0
E. coli 55989	38	22	70	27	0.2887	0.6662	0
E. coli ABU 83972	40	24	65	29	0.3964	0.7436	0
E. coli B str. REL606	37	23	71	26	0.1569	0.5891	0
E. coli BL21-Gold(DE3)pLysS AG	38	21	75	28	0.2891	0.6756	0
E. coli BL21(DE3) [iECD1391]	32	17	80	33	0.5789	0.7765	0
E. coli str. K-12 DH10B	39	21	69	28	0.4795	0.7536	0
E. coli ED1a	38	23	68	28	0.2966	0.6803	0
E. coli O157:H7	38	25	70	24	0.0788	0.5211	0
E. coli HS	35	24	67	30	0.2288	0.653	0
E. coli IAI1	40	23	67	27	0.3824	0.7008	0
E. coli IAI39	37	22	71	27	0.2174	0.6396	0
E. coli NA114	38	22	70	27	0.2887	0.6662	0
E. coli O103:H2 str. 12009	37	24	67	30	0.3043	0.6903	0
E. coli O111:H- str. 11128	39	24	68	26	0.2211	0.6213	0
E. coli O26:H11 str. 11368	40	23	64	31	0.7321	0.8424	0
E. coli IHE3034	38	21	68	30	0.5982	0.7983	0
E. coli ATCC 8739	36	24	69	29	0.2244	0.6304	0
E. coli 536	37	22	71	27	0.2174	0.6396	0
E. coli O157:H7 str. Sakai	49	27	95	37	0.2777	0.7068	0
E. coli S88	42	21	67	27	0.5975	0.806	0
E. coli SE11	39	23	66	29	0.488	0.7451	0
E. coli SMS-3-5	39	20	70	27	0.4735	0.7521	0
E. coli O157:H7 str. TW14359	39	23	68	26	0.223	0.6483	0
E. coli UMN026	38	22	70	27	0.2887	0.6662	0
E. coli W [iECW_1372]	38	22	68	29	0.3868	0.7366	0
E. coli O55:H7 str. CB9615	36	23	69	30	0.298	0.6805	0
Helicobacter pylori 26695	36	21	70	31	0.4822	0.7592	0
Synechocystis sp. PCC 6803	39	24	70	30	0.3086	0.6964	0
E. coli K-12 MG1655 [iJO1366]	44	23	64	31	0.8664	0.9266	0
E. coli K-12 MG1655 [iJR904]	36	23	69	29	0.2935	0.6578	0
E. coli LF82	40	23	68	26	0.2923	0.665	0
Thermotoga maritima MSB8	38	24	67	29	0.3027	0.6853	0
Shigella flexneri 2a str. 2457T	19	9	62	28	0.99	0.9534	0
Shigella boydii CDC 3083-94	32	28	47	31	0.4883	0.7538	0
Methanosarcina barkeri str. Fusaro	18	43	21	51	1	1.0166	0

Table S25: HGT-acquired strictly essential metabolic genes and operons. Each row corresponds to a given species or strain. Columns, from left to right, show species (strain) names, the number of strictly essential metabolic genes that are neither HGT-acquired nor belong to an operon $(--)$, the number of strictly essential metabolic genes that are not HGT-acquired but do belong to an operon $(-+)$, the number of strictly essential metabolic genes that are HGT-acquired, but do not belong to an operon $(+-)$, the number of strictly essential metabolic genes that are both HGT-acquired and belong to an operon $(++)$, the P value of a Fisher exact test on this data, the odds ratio (defined as the odds of being HGT-acquired for operonic (strictly essential) metabolic genes divided by the odds of being HGT-aquired for non-operonic (strictly essential) metabolic genes), and whether the null hypothesis of a lack of association between HGT and being on an operon (among strictly essential metabolic) is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). We do not observe a significant association between horizontally transferred genes and genes belonging to operons s (among strictly essential genes) in any genome.

Species (strain)	(--)	(-+)	(+-)	(++)	P-value	Odds ratio	Hypothesis rejected
Methanosarcina barkeri str. Fusaro	18	43	21	51	1	1.0166	0
Geobacter metallireducens GS-15	57	22	117	61	0.3858	1.3508	0
E. coli BL21(DE3) [iB21 1397]	75	20	164	57	0.3948	1.3034	0
E. coli BW2952	71	31	168	85	0.6177	1.1588	0
E. coli CFT073	35	20	78	60	0.4198	1.3462	0
E. coli O127:H6	58	26	142	76	0.5877	1.1939	0
E. coli 042	68	32	148	69	1	0.9907	0
E. coli 55989	67	30	146	74	0.6977	1.132	0
E. coli ABU 83972	62	30	145	72	0.99	1.0262	0
E. coli B str. REL606	67	31	145	73	0.7965	1.0881	0
E. coli BL21-Gold(DE3)pLysS AG	67	30	146	73	0.6984	1.1167	0
E. coli BL21(DE3) [iECD1391]	62	24	152	79	0.3454	1.3427	0
E. coli DH1 [iEcDH1 1363]	64	31	145	73	1	1.0394	0
E. coli O157:H7	64	29	145	75	0.6941	1.1415	0
E. coli HS	67	33	148	69	0.8972	0.9466	0
E. coli NA114	61	30	142	77	0.7933	1.1026	0
E. coli O103:H2 str. 12009	68	32	138	69	0.897	1.0625	0
E. coli O111:H- str. 11128	66	30	140	71	0.6967	1.1157	0
E. coli O26:H11 str. 11368	64	30	144	75	0.7941	1.1111	0
E. coli IHE3034	60	32	142	75	0.99	0.9903	0
E. coli ATCC 8739	68	32	141	72	0.7977	1.0851	0
E. coli 536	62	29	144	74	0.7917	1.0987	0
E. coli O157:H7 str. Sakai	65	28	138	76	0.4312	1.2785	0
E. coli S88	60	30	145	73	1	1.0069	0
E. coli SE11	66	30	145	71	0.7949	1.0772	0
E. coli SE15	75	34	176	80	1	1.0027	0
E. coli SMS-3-5	73	28	149	68	0.5999	1.1898	0
E. coli O157:H7 str. TW14359	65	30	138	76	0.5193	1.1932	0
E. coli W [iECW_1372]	68	28	143	71	0.5125	1.2058	0
E. coli KO11FL	67	31	143	71	0.8966	1.0731	0
E. coli ETEC H10407	65	30	146	68	0.99	1.0091	0
E. coli O55:H7 str. CB9615	68	29	152	72	0.7936	1.1107	0
E. coli LF82	60	29	145	72	1	1.0273	0
E. coli O83:H1 str. NRG 857C	59	28	146	73	0.8934	1.0536	0
Shigella flexneri 2a str. 2457T	58	30	127	78	0.5976	1.1874	0
Shigella flexneri 5 str. 8401	61	31	117	76	0.3638	1.2782	0
E. coli UM146	59	30	147	72	0.8944	0.9633	0
E. coli UMNK88	70	32	137	69	0.7967	1.1017	0
E. coli UTI89	63	30	141	75	0.6967	1.117	0
Bacillus subtilis str. 168	39	13	147	63	0.6088	1.2857	0

Table S26: HGT-acquired conditionally essential metabolic genes and operons. Each row corresponds to a given species or strain. Columns, from left to right, show species (strain) names, the number of conditionally essential metabolic genes that are neither HGT-acquired nor belong to an operon ($--$), the number of conditionally essential metabolic genes that are not HGT-acquired but do belong to an operon ($-+$), the number of conditionally essential metabolic genes that are HGT-acquired, but do not belong to an operon ($+-$), the number of conditionally essential metabolic genes that are both HGT-acquired and belong to an operon ($++$), the P value of a Fisher exact test on this data, the odds ratio (defined as the odds of being HGT-acquired for operonic (conditionally essential) metabolic genes divided by the odds of being HGT-aquired for non-operonic (conditionally essential) metabolic genes), and whether the null hypothesis of a lack of association between HGT and being on an operon (among conditionally essential metabolic) is rejected (1) or not (0). The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). We do not observe a significant association between horizontally transferred genes and genes belonging to operons (among conditionally essential genes) in any genome.

Species(strain)	Hypothesis rejected	P-value	Kuiper test statistic	Critical value
Methanosarcina barkeri str. Fusaro	0	0.3153	0.0851	0.1074
Geobacter metallireducens GS-15	0	0.9869	0.047	0.1066
E. coli BL21(DE3) [iB21 1397]	0	0.1343	0.0939	0.1001
E. coli BW2952	0	0.1729	0.0895	0.0999
E. coli CFT073	0	0.0825	0.107	0.1064
E. coli O127:H6	1	0.044	0.112	0.1032
E. coli 042	0	0.1515	0.0891	0.0972
E. coli 55989	0	0.0715	0.1029	0.0994
E. coli ABU 83972	0	0.3153	0.0774	0.0977
E. coli B str. REL606	0	0.0765	0.1014	0.0996
E. coli BL21-Gold(DE3)pLysS AG	0	0.085	0.0995	0.0994
E. coli BL21(DE3) [iECD1391]	1	0.0353	0.1215	0.1091
E. coli DH1 [iEcDH1 1363]	0	0.0866	0.0989	0.0993
E. coli O157:H7	0	0.118	0.0971	0.1012
E. coli HS	0	0.2239	0.0835	0.0985
E. coli NA114	0	0.4126	0.0739	0.0991
E. coli O103:H2 str. 12009	0	0.1342	0.0925	0.0983
E. coli O111:H- str. 11128	0	0.1515	0.0902	0.0985
E. coli O26:H11 str. 11368	0	0.0931	0.0981	0.0994
E. coli IHE3034	1	0.073	0.1043	0.1014
E. coli ATCC 8739	0	0.1765	0.0868	0.0974
E. coli 536	0	0.4988	0.069	0.0977
E. coli O157:H7 str. Sakai	1	0.0398	0.1121	0.1019
E. coli S88	0	0.0761	0.1047	0.1025
E. coli SE11	0	0.0698	0.102	0.098
E. coli SE15	1	0.0329	0.1126	0.1001
E. coli SMS-3-5	0	0.1463	0.0888	0.096
E. coli O157:H7 str. TW14359	0	0.1343	0.0931	0.0994
E. coli W [iECW_1372]	1	0.0353	0.1094	0.0981
E. coli KO11FL	0	0.2106	0.0854	0.0988
E. coli ETEC H10407	0	0.0765	0.0987	0.0972
E. coli O55:H7 str. CB9615	0	0.4512	0.0709	0.0975
E. coli LF82	0	0.2106	0.0877	0.1019
E. coli O83:H1 str. NRG 857C	0	0.2106	0.0872	0.1016
Shigella flexneri 2a str. 2457T	0	0.2604	0.0826	0.0999
Staphylococcus aureus N315	0	0.2106	0.6611	0.768
Shigella flexneri 5 str. 8401	0	0.1185	0.092	0.0962
E. coli UM146	0	0.0718	0.106	0.1026
E. coli UMNK88	0	0.4055	0.0707	0.0942
E. coli UTI89	0	0.1703	0.0904	0.1004
Klebsiella pneumoniae MGH78578	0	0.2106	0.0831	0.096
Bacillus subtilis str. 168	0	0.8512	0.0705	0.1249

Table S27: Clustering of HGT-acquired metabolic genes (including both essential and non-essential ones). Each row corresponds to a bacterial species or strain. Columns, from left to right, show species (strain) name, whether the null hypothesis of uniform distribution of metabolic genes acquired by HGT (including both essential and non-essential ones) is rejected by Kuiper's test (1) or not (0), the P-value of the test, Kuiper's test statistics, and the critical value of this statistic above which the null hypothesis is rejected. The P-values are adjusted for multiple-testing using the Benjamini-Hochberg correction (32). Only in 6 of the 55 genomes the null hypothesis is rejected, i.e., conditionally essential operons are significantly clustered.

References:

1. van Nimwegen E, Crutchfield JP, Huynen M (1999) Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A* 96(17):9716–20.
2. Sung W, et al. (2016) Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. *G3: Genes|Genomes|Genetics* 6(8):2583–2591.
3. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17(10):589–96.
4. Nilsson AI, et al. (2005) Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A* 102(34):12112–6.
5. Sung W, et al. (2015) Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation-Accumulation Experiments. *Mol Biol Evol* 32(7):1672–83.
6. Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc Natl Acad Sci* 109(41):E2774–E2783.
7. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A* 109(45):18488–92.
8. Sung W, et al. (2012) Extraordinary genome stability in the ciliate Paramecium tetraurelia. *Proc Natl Acad Sci* 109(47):19339–19344.
9. Pál C, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440(7084):667–670.
10. Chan CX, Beiko RG, Darling AE, Ragan MA (2010) Lateral Transfer of Genes and Gene Fragments in Prokaryotes. *Genome Biol Evol* 1(0):429–438.
11. Dilthey A, Lercher MJ (2015) Horizontally transferred genes cluster spatially and metabolically. *Biol Direct* 10:72.
12. Igarashi N, et al. (2001) Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J Mol Evol* 52(4):333–41.
13. Omelchenko M V, Makarova KS, Wolf YI, Rogozin IB, Koonin E V (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol* 4(9):R55.
14. Lin CH, Bourque G, Tan P (2008) A comparative synteny map of Burkholderia species links large-scale genome rearrangements to fine-scale nucleotide variation in prokaryotes. *Mol Biol Evol* 25(3):549–58.
15. Itoh T, Takemoto K, Mori H, Gojobori T (1999) Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol Biol Evol* 16(3):332–46.
16. de Daruvar A, Collado-Vides J, Valencia A (2002) Analysis of the Cellular Functions of Escherichia coli Operons and Their Conservation in Bacillus subtilis. *J Mol Evol* 55(2):211–221.
17. Overbeek R, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33(17):5691–702.
18. Dan Tenenbaum (2018) KEGGREST: Client-side REST access to KEGG. *R Packag version 1181*. Available at: https://bioconductor.org/packages/release/bioc/html/KEGGREST.html.
19. King ZA, et al. (2015) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*:gkv1049-.
20. Lathe WC, Snel B, Bork P (2000) Gene context conservation of a higher order than operons. *Trends Biochem Sci* 25(10):474–9.
21. Lawrence JG, Roth JR (1996) Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics* 143(4):1843–1860.
22. Pál C, Hurst LD (2004) Evidence against the selfish operon theory. *Trends Genet* 20(6):232–234.
23. Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* 15(6):809–19.
24. Mao F, Dam P, Chou J, Olman V, Xu Y (2009) DOOR: a database for prokaryotic

operons. *Nucleic Acids Res* 37(Database):D459–D463.
25. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245–8.
26. Feist AM, et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3:121.
27. Lercher MJ, Pál C (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol* 25(3):559–67.
28. Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420(6912):186–9.
29. Vieira-Silva S, Rocha EPC (2010) The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet* 6(1):e1000808.
30. Fong SS, Palsson BØ (2004) Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. *Nat Genet* 36(10):1056–8.
31. Fong SS, Marciniak JY, Palsson BO (2003) Description and Interpretation of Adaptive Evolution of Escherichia coli K-12 MG1655 by Using a Genome-Scale In Silico Metabolic Model. *J Bacteriol* 185(21):6400–6408.
32. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57:289–300.